University of Alberta

Transposon Mutagenesis and Characterization of Clavam Biosynthetic Genes in *Streptomyces clavuligerus*

by

Hui Cai ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Science

in

Microbiology and Biotechnology

Department of Biological Sciences

Edmonton, Alberta

Fall, 2003

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisisitons et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-612-87729-9
Our file Notre référence
ISBN: 0-612-87729-9

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou aturement reproduits sans son autorisation.

Canada

University of Alberta

Library Release Form

Name of Author: Hui Cai

Title of Thesis: Transposon Mutagenesis and Characterization of Clavam Biosynthetic Genes in *Streptomyces clavuligerus*

Degree: Master of Science

Year this Degree Granted: 2003

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly, or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

Date: August 27, 2003

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled Transposon Mutagenesis and Characterization of Clavam Biosynthetic Genes in *Streptomyces clavuligerus* submitted by Hui Cai in partial fulfillment of the requirements for the degree of Master of Science in Microbiology and Biotechnology.

Date approved: August 26, 2003

Abstract

An approximately 11 kb genetic region located next to the *ceaS1* gene of the paralogue gene cluster in *Streptomyces clavuligerus* was investigated by a Tn*5*-based *in vitro* transposon mutagenesis system. The cosmid clone 6G9 that contained the 11 kb fragment was used as the target for the *in vitro* transposition reaction. The reaction generated >1000 mutant clones in *Escherichia coli*. Over 150 mutant clones were analyzed by restriction digestion, and 30 of them were determined to carry Tn*5* on the 11 kb fragment. The transposon-flanking regions were then sequenced bi-directionally by using the transposon-derived primers, thereby determining the target sites and orientation of the transposon insertions. Eleven putative open reading frames (ORFs) were identified. One or two mutant clones that lie in each of the ORFs were introduced into wildtype *S. clavuligerus* via conjugation to create mutant strains. The involvement of the disrupted genes in clavam biosynthesis was assessed by bioassay and High Performance Liquid Chromatography.

Acknowledgement

I would like to express my sincere appreciation to my supervisor Dr. Susan E. Jensen for her guidance, patience and continued support during the completion of this thesis.

I would also like to express my gratitude to my supervisory committee members, Dr. Brenda K. Leskiw and Dr. Jonathan Dennis, for their guidance and suggestions towards my research project. Special thanks to Annie Wong and Cecilia Anders, who gave me help and advice on my research project, especially their technical assistance on HPLC. Also thanks to the other members of my laboratory, Dr. Liru Wang, Wei Wu, Kapli Tahlan and Jingru Li, for their help, suggestions and friendship.

Finally, I would like to thank my parents, brother and my boyfriend Wei Zhang, for their support over the many years.

TABLE OF CONTENTS

	Page Number
I. INTRODUCTION	1
II. MATERIALS AND METHODS	55
II.1 Materials	55
II.2 Bacterial strains, growth media, culture conditions, plasmids and cosmids	56
II.2.1 Bacterial strains, growth media and culture conditions	56
II.2.2 Plasmids and cosmids	60
II.3 Isolation of DNA	63
II.3.1 Isolation of plasmid and cosmid DNA from *E. coli*	63
II.3.2 Preparation of genomic DNA from *Streptomyces*	63
II.3.3 Determining DNA Concentration and Evaluating DNA Quality	64
II.4 Introduction of DNA into bacterial cells	65
II.4.1 Transformation of DNA into *E. coli*	65
II.4.2 Conjugation of DNA into *S. clavuligerus*	66
II.5 Techniques used for DNA analysis	67
II.5.1 Agarose gel electrophoresis	67
II.5.2 Purification of DNA fragments from agarose gels	67
II.5.3 Southern hybridization	68
II.5.4 DNA sequencing and sequence analysis	70

II.6 Generation of DNA probe fragments by PCR 71

II.7 Construction of *S. clavuligerus* mutants 72

II.7.1 *In vitro* transposon mutagenesis 72

II.7.2 Construction of *S. clavuligerus* mutant strains 73

II.8 β-lactam bioassay and HPLC analysis 76

II.8.1 β-lactam bioassay 77

II.8.2 Analysis of clavulanic acid and other clavam metabolites by HPLC 78

II.8.2.1 Analysis of clavams in liquid cultures 78

II.8.2.2 Analysis of clavams in solid cultures 79

III. RESULTS 80

III.1 Identifying the *Eco*RI fragment of interest on cosmid 6G9 80

III.2 *In vitro* transposon mutagenesis 87

III.3 Analysis of transposon-containing clones in *E. coli* 88

III.3.1 Screening of transposon-containing clones 88

III.3.2 Sequence analysis of the 11 kb *Eco*RI fragment on cosmid 6G9 93

III.3.3 Deduced gene products of identified open reading frames 116

III.3.3.1 The *orfB* gene and its deduced product 117

III.3.3.2 The *orfC* gene and its deduced product 122

III.3.3.3 The *orfD* gene and its deduced product 124

III.3.3.4 The *orfE* gene and its deduced product 126

III.3.3.5 The *orfF* gene and its deduced product 127

III.3.3.6 The *orfG* gene and its deduced product 129

III.3.3.7 The *orfH* gene and its deduced product 129

III.3.3.8 The *orfI* gene and its deduced product 131

III.3.3.9 The *orfJ* gene and its deduced product 133

III.3.3.10 The *orfK* gene and its deduced product 134

III.3.3.11 The *orfL* gene and its deduced product 135

III.3.4 Target sites of the transposition reaction 136

III.4 Construction of *S. clavuligerus* mutants 138

III.5 Analysis of the *S. clavuligerus* mutants 143

III.5.1 Qualitative analysis of clavam biosynthesis in *S. clavuligerus* mutants 144

III.5.2 Quantitative analysis of clavam biosynthesis in *S. clavuligerus* mutants 151

III.5.2.1 β-lactam antibiotic bioassay 151

III.5.2.2 HPLC analysis 155

III.6 Confirmation of the *S. clavuligerus* mutants by Southern analysis 160

IV. DISCUSSION 169

V. REFERENCES 190

LIST OF TABLES

	Page Number
Table 1. List of bacterial strains used in this study.	57
Table 2. Frequencies of transposon insertion into the different *Eco*RI fragments of cosmid 6G9.	92
Table 3. Oligonucleotide primers designed from new sequence data acquired.	94
Table 4. Properties of the newly discovered open reading frames.	113
Table 5. Deduced gene products of the ORFs present on the 11 kb *Eco*RI fragment on cosmid 6G9 and their homologous proteins.	118
Table 6. Tn*5* insertion sites.	137
Table 7. Calculated transposon insertion target data.	139
Table 8. Qualitative analysis of clavam biosynthesis of *S. clavuligerus* wildtype and mutants.	149
Table 9. β-lactam bioassay results for *S. clavuligerus* wildtype and mutants.	153
Table 10. HPLC analysis of clavulanic acid and other clavam metabolites production by *S. clavuligerus* wildtype and mutants on Soy medium (I).	156
Table 11. HPLC analysis of clavulanic acid and other clavam metabolites production by *S. clavuligerus* wildtype and mutants on Soy medium (II).	161

LIST OF FIGURES

	Page Number
Figure 1. *β*-Lactam products of *S. clavuligerus.*	4
Figure 2. Biosynthetic pathway leading to clavulanic acid and the other clavams in *S. clavuligerus.*	8
Figure 3. Arrangement of clavulanic acid biosynthetic gene cluster in *S. clavuligerus.*	10
Figure 4. Arrangement of the clavulanic acid paralogue biosynthetic gene cluster in *S. clavuligerus.*	13
Figure 5. Arrangement of the *5S* clavam biosynthetic gene cluster in *S. clavuligerus.*	26
Figure 6. Proposed pathway for the biosynthesis of clavulanic acid and the *5S* clavams beginning at proclavaminic acid.	29
Figure 7. Structures of bacterial transposable elements.	35
Figure 8. Mechanism of conservative transposition.	38
Figure 9. Mechanism of replicative transposition.	41
Figure 10. Tn*5* structure.	51
Figure 11. Map of modified Tn*5* transposon-containing plasmid pQM5062.	61
Figure 12. *In vitro* transposon mutagenesis reaction.	74
Figure 13. Partial map of cosmid 6G9.	82
Figure 14. Southern hybridization analysis of *Eco*RI fragments of cosmid 6G9.	85
Figure 15. Determination of the transposon-containing *Eco*RI fragments by restriction digestion.	89

Figure 16. Sequence data obtained shown in original contigs. 96

Figure 17. Compiled 11.4 kb nucleotide sequence data. 98

Figure 18. Prediction of the open reading frames present on the 11.4 kb DNA fragment by FramePlot 3.0 beta. 110

Figure 19. Arrangement of the newly discovered open reading frames relative to the sites and orientations of transposon insertions. 114

Figure 20. Examples of HPLC-chromatograms of "freeze-squeeze" samples by *S. clavuligerus* wildtype and mutant strains on Soy solid medium. 145

Figure 21. Examples of HPLC chromatograms of clavulanic acid and clavam metabolites production by *S. clavuligerus* wildtype and mutant strains on Soy medium (I). 157

Figure 22. Examples of HPLC chromatograms of clavulanic acid and clavam metabolites production by *S. clavuligerus* wildtype and mutant strains on Soy medium (II). 162

Figure 23. Southern hybridization analysis of genomic DNA from wildtype *S. clavuligerus* and 27 *S. clavuligerus* mutants disrupted with Tn*5*. 166

LIST OF ABBREVIATIONS

Apra	Apramycin
CAS	Clavaminate synthase
CDD	Conserved domain database
Cm	Chloramphenicol
DR	Direct repeat
EGFP	Enhanced green fluorescent protein
egfp	Gene encoding enhanced green fluorescent protein
HPLC	High performance liquid chromatography
HTH	Helix-turn-helix
IE	Inside end
Inh	Transposase inhibitor
IR	Inverted repeat
Km	Kanamycin
MFS	Major facilitator superfamily
OE	Outside end
ORF	Open reading frame
*ori*T	Origin of transfer
PBP	Penicillin-binding proteins
PLP	Pyridoxal phosphate
SA	Starch asparagine
SARP	*Streptomyces* antibiotic regulatory proteins
Tnp	Transposase
TSB	Trypticase soy broth
TSBS	TSB + 1% starch
v/v	Volume/volume
w/v	Weight/volume

I. INTRODUCTION

β-lactam antibiotics are modified peptides that act on peptidoglycan-related enzymes in prokaryotes, affecting the cell-wall integrity. Since Fleming's accidental discovery of a *Penicillium* contaminating isolate and its remarkable antibacterial properties over 70 years ago, various β-lactam antibiotics have been discovered and developed. These include penicillins produced by *Penicillium chrysogenum*, cephalosporins formed by *Cephalosporium acremonium*, cephamycins, clavams and carbapenems produced by actinomycetes, and monocyclic β-lactams produced by actinomycetes and unicellular bacteria. The discovery and development of β-lactam antibiotics are among the most successful examples of natural product application and chemotherapy.

Individual and collaborative research groups from industry and academia have conducted studies on the microbiological, biochemical, genetic and chemical aspects of β-lactam compounds for many years. During the early years after the discovery of β-lactams, research was focused on widespread screening for novel antibiotics and their producer organisms, increasing antibiotic production by random mutagenesis, and selection of strains with increased production yields. Much of the information gained at this time had little to do with the genetic characterization of the genes and enzymes involved in the biosynthetic pathway.

Within the past three decades, intensive research has been undertaken to identify and analyze structural and regulatory biosynthetic genes responsible for enzymatic synthesis, and to characterize the enzymes catalyzing the biosynthetic pathways and limiting the

production of β-lactams. The research has resulted in a far greater understanding of the synthesis and mechanisms of action of β-lactams, and the improvement of β-lactam compounds with respect to potency, breadth of spectrum, activity against resistant pathogens, stability and pharmacokinetic properties. The role of β-lactam antibiotics has remained prominent. Today, they account for about 60% of the $23 billion worldwide market for antibiotics (Demain 1999).

Although several hundreds of compounds with antibiotic activity have been isolated from microorganisms over the years, only a few of them, which are highly effective against microorganisms but have minimal toxicity to humans, are clinically useful. For this reason, actinomycetes stand out as preeminent producers of therapeutically valuable compounds. Among this group, the genus *Streptomyces* is responsible for the production of more than 70% of our clinically useful antibiotics.

Streptomyces spp. are Gram-positive, filamentous soil bacteria belonging to the order *Actinomycetales*. *Streptomyces* are known for their production of a large variety of biologically active secondary metabolites, degradative enzymes and enzyme inhibitors, and for their complex life cycle. The life cycle starts with the germination of a spore, giving rise to a branching network of multinucleoid hyphae known as substrate mycelium. Morphological differentiation begins with the formation of aerial hyphae, which grow into the air away from the surface of the colony. Then these aerial hyphae undergo septation into uninucleoid compartments and metamorphose into chains of spores. Antibiotic production usually occurs during the late stages of the development of aerial

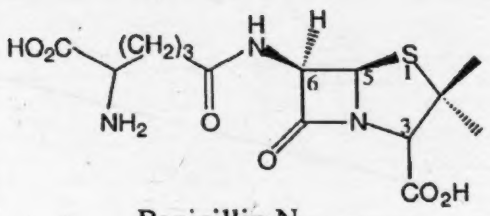
mycelium on solid medium or just before the commencement of stationary phase in liquid cultures. *Streptomyces* are characterized as having a G+C content of 69-78%, and an unusual linear chromosome as well as both linear and circular plasmids. Genes encoding secondary metabolites, especially antibiotic biosynthetic pathways, are usually clustered on *Streptomyces* chromosomes (Kieser et al. 2000).

Within the genus, *Streptomyces clavuligerus* has been the subject of extensive study in the past 30 years because of its ability to produce β-lactam metabolites with antibiotic, antifungal and β-lactamase-inhibitory activities. *S. clavuligerus* was isolated in 1971 during a screen for producers of β-lactam compounds with improved resistance to β-lactamases (Nagarajan 1971). Now it is known to produce at least seven different β-lactam compounds as well as two structurally unrelated antibiotics, holomycin and tunicamycin. The major β-lactam compounds produced by *S. clavuligerus* are shown in Figure 1. Penicillin N and cephamycin C are classical antibiotics while clavam-2-carboxylate, hydroxymethyl clavam, alanylclavam and 2-formyloxymethyl clavam possess weak antifungal activity, and alanylclavam also has antibacterial activities (Pruess and Kellett 1983).

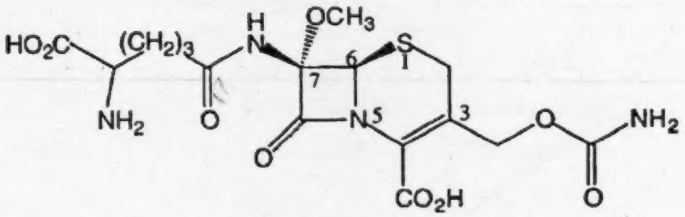
Besides, *S. clavuligerus* is well known as the species used for the industrial production of clavulanic acid. Clavulanic acid is a poor antibiotic, but a powerful suicide inhibitor of a large number of β-lactamases (Baggaley et al. 1997). It is used clinically in combination with other antibiotics, such as amoxycillin and ticarcillin, to overcome acquired resistance of antibiotic resistant organisms. This has made clavulanic acid a

Figure 1. β-Lactam products of *S. clavuligerus* (Thai et al. 2001). (a) Conventional β-lactam antibiotics. (b) Clavam metabolites.

(a)

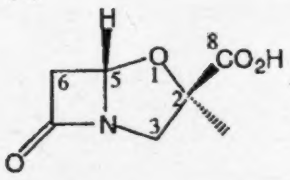


Penicillin N

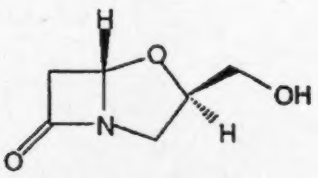


Cephamycin C

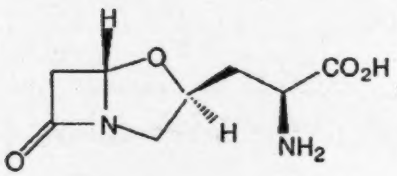
(b)



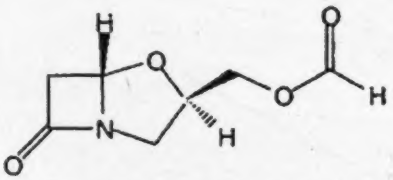
Clavam-2-carboxylate



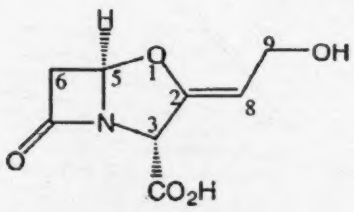
Hydroxymethyl clavam



Alanyl clavam



Formyl-oxymethyl clavam



Clavulanic acid (C-5R)

product valued in excess of a billion dollars/annum and makes it important to understand the biochemical and molecular genetic features involved in the production of clavulanic acid.

In structure, clavulanic acid differs from the other four clavams in that it has C-*5R* stereochemistry, which is opposite to the C-*5S* stereochemistry of the other clavams. Furthermore, it has a carboxyl group at the C-3 position not seen in the other clavams. The only clavam metabolite other than clavulanic acid known to show β-lactam inhibitory activity is the late biosynthetic intermediate clavulanate-9-aldehyde which also has the *3R, 5R* configuration. Therefore, β-lactam inhibitory activity seems to be associated with the *3R, 5R* stereochemistry since the other clavam metabolites which have a *3S, 5S* configuration lack this activity. These other clavam metabolites will be referred to as the *5S* clavams for the rest of this work.

Genes encoding enzymes for antibiotic biosynthesis are clustered on the chromosomes of bacterial antibiotics producers, and *S. clavuligerus* is no exception. Sequence analysis, hybridization studies and insertional inactivation experiments indicated that genes responsible for clavulanic acid and cephamycin C production in *S. clavuligerus* are organized into two separate clusters, but are found to be located adjacent to one another on the chromosome and form a "super-cluster" (Ward and Hodgson 1993). The branched biosynthetic pathway leading to cephamycins, penicillins and cephalosporins has been well characterized. Many biosynthetic genes have been cloned and sequenced, and their encoded pathway enzymes have been purified and characterized.

Excellent reviews on genes and enzymes involved in penicillin and cephamycin biosynthesis (Martin 1998, Liras 1999, Demain and Elander 1999) can be consulted.

While the biosynthetic pathway for penicillin and cephamycin is known and most of the biosynthetic genes have been cloned and sequenced, the understanding of the clavulanic acid biosynthetic pathway is still at an early stage of development. The biosynthesis of clavulanic acid (Figure 2) in *S. clavuligerus* begins with a condensation reaction of arginine with 3-phosphoglyceraldehyde to form N^2-(2-carboxyethyl)-arginine (CEA) (Khaleeli et al. 1999). This biosynthetic step is catalyzed by a TPP-dependent carboxyethylarginine synthase (CEAS) encoded by the *ceaS* (also named *orf2*) gene (Perez-Redondo et al. 1999) in the clavulanic acid biosynthetic gene cluster (Figure 3). The synthesis of the β-lactam ring from CEA is catalyzed by β-lactam synthetase (BLS), encoded by the *bls* gene. This process formed the monocyclic β-lactam deoxyguanidinoproclavaminic acid (DGPC) (Bachmann et al. 1998; Bachmann and Townsend 2000). DGPC is then hydroxylated by the multifunctional enzyme clavaminate synthase (CAS) to form guanidinoproclavaminic acid. The enzyme proclavaminate amidinohydrolase (PAH), encoded by the *pah* gene in the cluster, then hydrolyzes the guanidino group residue of the guanidinoproclavaminic acid to form proclavaminic acid (Wu et al. 1995). This is followed by the two step CAS-mediated oxidative cyclization of proclavaminic acid to form clavaminic acid (Elson et al. 1987). The final steps of the biosynthetic pathway are poorly understood. The enzyme clavulanate-9-aldehyde

Figure 2. Biosynthetic pathway leading to clavulanic acid and the other clavams in *S. clavuligerus* (Jensen et al. 2000; Jensen et al. 2003a). The enzymes and the corresponding genes at each step are indicated.

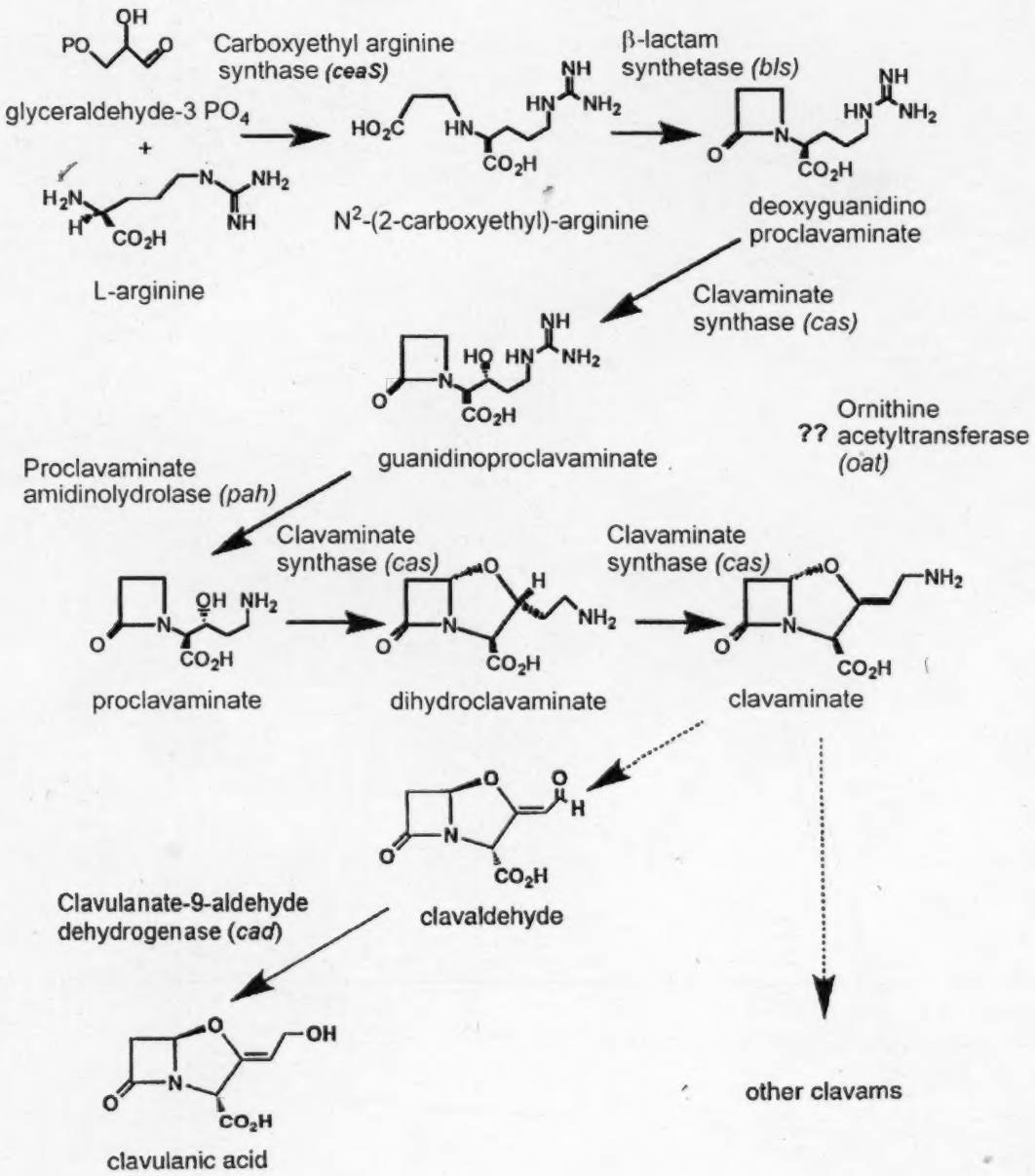
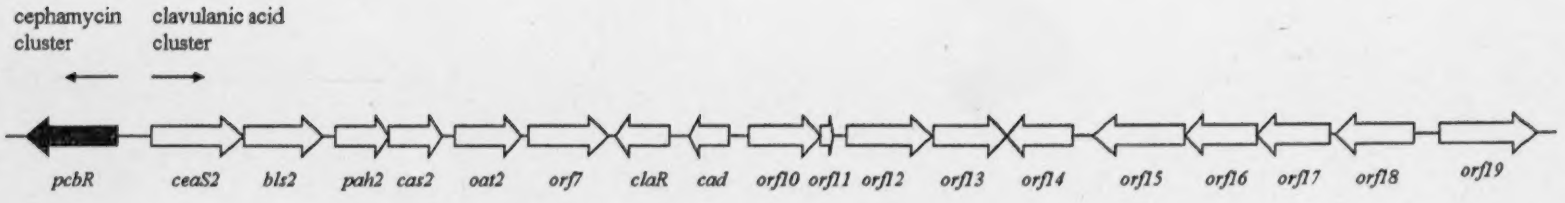

glyceraldehyde-3 PO4
+
L-arginine
Carboxyethyl arginine synthase (ceaS)
N2-(2-carboxyethyl)-arginine
β-lactam synthetase (bls)
deoxyguanidino proclavaminate
Clavaminate synthase (cas)
guanidinoproclavaminate
?? Ornithine acetyltransferase (oat)
Proclavaminate amidinolydrolase (pah)
proclavaminate
Clavaminate synthase (cas)
dihydroclavaminate
Clavaminate synthase (cas)
clavaminate
clavaldehyde
Clavulanate-9-aldehyde dehydrogenase (cad)
clavulanic acid
other clavams

Figure 3. Arrangement of clavulanic acid biosynthetic gene cluster in *S. clavuligerus* (Jensen et al. 2000; Jensen et al. 2003a). The line in the middle represents the *S. clavuligerus* chromosome, and the arrows on the line represent the open reading frames and the direction of their transcription. White arrows represent clavulanic acid biosynthetic genes; the grey arrow represents a cephamycin biosynthetic gene.

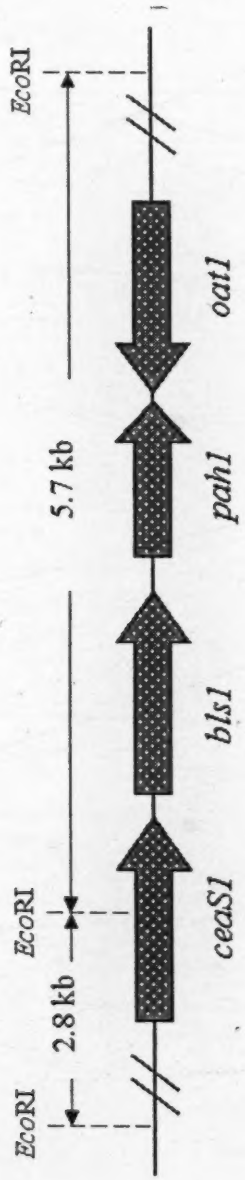

cephamycin cluster
clavulanic acid cluster
pcbR
ceaS2
bls2
pah2
cas2
oat2
orf7
claR
cad
orf10
orf11
orf12
orf13
orf14
orf15
orf16
orf17
orf18
orf19

reductase (CAR) is likely to be involved in the direct conversion of the highly reactive intermediate clavulanate-9-aldehyde into clavulanic acid (Nicholson et al. 1994). CAR is encoded by the *car* (also named *cad*) gene which is present in the clavulanic acid biosynthetic gene cluster (Perez-Redondo et al. 1998). However, the reaction(s) leading to the formation of the clavulanate-9-aldehyde from the clavaminate acid are unknown. Similarly, the reactions leading to the formation of *5S* clavams are unknown.

The steps leading to the formation of clavaminic acid are considered as the early steps in the biosynthesis of clavulanic acid and *5S* clavams (Jensen et al. 2000). Duplicate copies, or paralogues, exist for each of the genes encoding early enzymes of the biosynthetic pathway. Recently, the paralogue genes of all the clavulanic acid early biosynthetic genes have been isolated. Since then, all of the early biosynthetic genes in the clavulanic acid biosynthesis gene cluster have been given the additional designation 2, and thus become *ceaS2*, *bls2*, *pah2*, *cas2* and *oat2*, to distinguish them from their paralogues, *ceaS1*, *bls1*, *pah1*, *cas1* and *oat1* (Tahlan et al. 2003). The newly found *ceaS1*, *bls1*, *pah1* and *oat1* genes form a paralogue gene cluster (Figure 4), which are located elsewhere on the chromosome of *S. clavuligerus*. Although the relative arrangement of *ceaS1*, *bls1* and *pah1* is similar to that of *ceaS2*, *bls2* and *pah2* in the clavulanic acid gene cluster, *cas1* is found elsewhere on the *S. clavuligerus* chromosome surrounded by the genes for *5S* clavam biosynthesis (Mosher et al. 1999). Therefore, a more complex evolution of the paralogue gene cluster must be involved.

One of the earliest developments in molecular genetics of clavulanic acid

Figure 4. Arrangement of the clavulanic acid paralogue biosynthetic gene cluster in *S. clavuligerus* (Tahlan et al. 2003). The line in the middle represents part of the cosmid 6G9, which contains a fragment of *S. clavuligerus* chromosomal DNA. The arrows on the line represent the open reading frames and the direction of their transcription. *Eco*RI sites are indicated.


EcoRI
5.7 kb
oat1
pah1
bls1
ceaS1
EcoRI
2.8 kb
EcoRI

biosynthesis came with the cloning of the *pah* gene which was originally thought to be associated with the cephamycin production, but subsequently found to be involved in clavulanic acid biosynthesis. This *pah* gene was renamed as *pah2* after isolation of its paralogues gene, *pah1* (Jensen et al. 2003b). The *pah* genes encode the enzyme proclavaminate amidinohydrolase (PAH), which is able to hydrolyse the guanidino group of the guanidinoproclavaminic acid to form proclavaminic acid (Elson et al. 1993; Wu et al. 1995). The *pah1* gene showed a high degree of similarity to *pah2* at both the nucleotide (72% identical) and amino acid (71% identical) levels. Mutants that are defective in *pah2* gene showed modest reduction in the production of clavulanic acid and no effect on the production of the *5S* clavam metabolites when grown on Soy medium, but complete loss of metabolite production in starch asparagine (SA) medium (Jensen et al. 2000). In contrast, mutants that are defective in *pah1* showed no significant effects on production of clavulanic acid or any of the *5S* clavam metabolites when grown on SA or Soy media. However double mutants with defects in both *pah1* and *pah2* were unable to produce clavulanic acid and all of the *5S* clavam metabolites in any of these media (Jensen et al. 2003b). This indicated that each of the paralogue genes could at least partially complement the mutation of the other, and thus disruption of only one copy of the paralogue genes could not completely block the production of clavams unless the second copy of the paralogues could also be knocked out.

Isolation and characterization of CAS was one of the first important advances in the understanding of clavulanic acid biosynthesis. Purification and characterization of CAS

showed that the enzyme was a mixture of two isoenzymes CAS1 and CAS2 with similar primary amino acid sequences and kinetic properties (Salowe et al. 1991). CAS1 and CAS2 were encoded by two separate genes *cas1* and *cas2* (Marsh et al. 1992). The alignment of *cas1* and *cas2* indicates that they are 87% identical at the nucleotide level. The homology of the two genes is more extensive between the central portions and more variable toward both termini (Marsh et a. 1992). The *cas2* gene is present within the clavulanic acid cluster immediately downstream of the *pah2* gene (Ward and Hodgson 1993), while *cas1* was found to be located in a cluster of genes for *5S* clavam biosynthesis more than 20 kb away on the chromosome of *S. clavuligerus* from *cas2* (Mosher et al. 1999). Disruption of *cas2* yielded a mutant unable to produce clavulanic acid in SA medium but still able to produce clavulanic acid in Soy medium (Paradkar and Jensen 1995). Disruption of *cas1* yielded a mutant that produced close to wildtype levels of clavulanic acid in SA medium but lower levels of both clavulanic acid and most of the *5S* clavams except for 2-alanylclavam on Soy media. However, a double mutant of *S. clavuligerus*, disrupted in both *cas1* and *cas2*, produced neither clavulanic acid nor any of the *5S* clavams, including 2-alanylclavam (Mosher et al. 1999).

The first gene in the clavulanic acid biosynthetic gene cluster adjacent to the cephamycin biosynthetic gene cluster is *ceaS2* (also named *orf2*). The *ceaS* genes, both *ceaS1 & 2*, encode enzyme carboxyethylarginine synthase. The *ceaS1* gene displays 73% identity at the nucleotide level and 66% identity at the amino acid level to the *ceaS2* gene. Mutants that are defective in either the *ceaS2* or *ceaS1* genes were completely blocked in

clavulanic acid and *5S* clavam metabolite biosynthesis in SA medium, and produced significantly reduced amount of clavulanic acid and *5S* clavam metabolites in Soy medium. However, the *ceaS1* and *ceaS2* double mutants were completely blocked in clavulanic acid and *5S* clavam production in any of these media (Tahlan et al. 2003).

The biosynthetic gene downstream to *ceaS2* in the clavulanic acid gene cluster in *S. clavuligerus* is *bls2*. The *bls* genes encode the enzyme β-lactam synthetase (Bachmann et al. 1998; Bachmann and Townsend, 2000). The *bls1* gene displays 60% identity at the nucleotide level and 49% identity at the amino acid level to the *bls2* gene. Similar to the situation seen with *pah1 & 2* and *ceaS1 & 2*, when *bls1* and *bls2* were knocked out individually, the mutants still retained some ability to produce clavulanic acid. However, when a *bls1/bls2* double mutant strain was prepared, no clavulanic acid or *5S* clavam production could be detected (Tahlan et al. 2003).

The *oat2* gene (also named *orf6*) in the clavulanic acid biosynthetic gene cluster in *S. clavuligerus* encodes a protein shown to possess ornithine acetyltransferase activity (Kershaw et al. 2002). The *oat1* gene displays 63% identity at the nucleotide level and 47% identity at the amino acid level to the *oat2* gene. Mutants that are defective in *oat2* were unable to produce clavulanic acid in SA medium, but produced approximately 40% of the wildtype level clavulanic acid in Soy medium (Jensen et al. 2000; Kershaw et al. 2002). Mutants that are defective in *oat1* produced clavulanic acid at the level that varied from 30 to 154% in SA medium and 57 to 145 % in Soy medium, as compared to the wildtype strain. The levels of *5S* clavam metabolites produced by both mutant types in

Soy medium were very variable. However, the *oat1/oat2* double mutant, unlike the double mutants of other paralogue genes, still produced some clavulanic acid and *5S* clavam metabolites (Tahlan et al. 2003). This may be attributed to the presence of another ornithine acetyltransferase enzyme, encoded by the *argJ* gene in the arginine biosynthetic gene cluster, in *S. clavuligerus* (Rodriguez-Garcia et al. 2000), which may compensate for the lack of OAT1 & 2 in the double mutants (Tahlan et al. 2003). Since OATs are normally involved in arginine biosynthesis, the role of OAT in the biosynthesis of clavams may be directed towards increasing intracellular concentrations of arginine for clavam biosynthesis rather than primary metabolism (Kershaw et al. 2002).

The final steps of the clavulanic acid pathway from clavaminic acid to clavulanic acid are poorly understood. It was therefore assumed that additional genes required for clavulanic acid production must exist. Since the genes involved in antibiotic biosynthesis are clustered in most actinomycetes, the DNA region downstream of *oat2* was sequenced and more genes were identified. The involvement of these ORFs in clavulanic acid biosynthesis was assessed by creating mutants with defects in each of these ORFs.

The *orf7* gene of the clavulanic acid biosynthetic gene cluster in *S. clavuligerus* encodes a protein that is similar to a group of peptide transport proteins (Hodgson et al. 1995). These proteins are the periplasmic binding protein components of transport systems. Mutants that are defective in *orf7* were unable to produce clavulanic acid even in Soy medium. However, the role of Orf7 in clavulanic acid biosynthesis is unclear.

Although the reactions involved in the final steps of the clavulanic acid pathway are

unclear, the existence of an aldehyde intermediate with the same stereochemistry as clavulanic acid has been demonstrated, and the presence of a NADPH-dependent dehydrogenase called clavulanate-9-aldehyde dehydrogenase (CAD) has been shown to reduce clavaldehyde into clavulanic acid (Nicholson et al. 1994). Both the size and the N-terminal amino acid sequence suggest that CAR is encoded by the *cad* (also named *car*) gene in the clavulanic acid gene cluster (Perez-Redondo et al. 1998). However, the rest of the final steps leading to formation of clavulanic acid remain unknown.

A regulatory gene, *claR*, is found in the clavulanic acid gene cluster. The gene is located downstream of, and in the same orientation as the *cad* gene in the cluster. *claR* encodes a 47 kD polypeptide with a significant degree of homology to many transcriptional activators of the LysR family. The protein resembles members of the LysR family in the presence of the characteristic helix-turn-helix (HTH) motifs, corresponding to the DNA binding regions of transcriptional activators. However, the placement of the HTH motif of ClaR differs from other LysR regulators. All the LysR proteins examined to date have the HTH motif present close to the N-terminus, but the HTH motifs of ClaR are located at both the N- and C-termini of the protein.

The ClaR protein seems specifically targeted at the regulation of clavulanic acid in *S. clavuligerus*. Disruption of *claR* yields a mutant that accumulates clavaminic acid but is unable to produce clavulanic acid (Paradkar et al. 1998; Perez-Redondo et al. 1998). In addition, in *claR*-disrupted mutants, transcripts of *orf7*, *cad* and *orf10* are not found, whereas expression of *ceaS2*, *bls2*, *pah2*, *cas2* and *oat2* can occur in the absence of ClaR.

Further analysis showed that individual disruption of *orf7*, *cad*, and *orf10* results in mutants that are defective in clavulanic acid production on both SA and Soy media. These results suggested that genes involved in the early stages of the clavulanic acid biosynthetic pathway, *ceaS2-oat2*, are independent of the regulation of *claR*, and genes *orf7-orf10* are involved in the final steps of the clavulanic acid biosynthetic pathway, and that their expression is regulated by *claR* (Paradkar et al. 1998).

Another transcriptional regulatory protein that controls the production of clavulanic acid is CcaR, which is encoded by the *ccaR* gene in the cephamycin biosynthetic gene cluster in *S. clavuligerus* (Perez-Llarena et al. 1997). CcaR belongs to the SARP (*Streptomyces* antibiotic regulatory proteins) family of proteins. It has been proposed that SARPs recognize specific heptameric repeats in SARP-regulated promoters (Wietzorrek and Bibb 1997). Disruption of *ccaR* resulted in a mutant that is unable to produce cephamycin C, clavulanic acid and *5S* clavams (Perez-Llarena et al. 1997, Alexander and Jensen 1998, Jensen et al. unpublished data). Although the mechanism of how CcaR regulates clavulanic acid production is unclear, it might be mediated by controlling the expression of *claR* in the clavulanic acid biosynthetic gene cluster. A *ccaR*-disrupted mutant did not express *claR*, which in turn, is found to control the transcription of late biosynthetic genes in clavulanic acid biosynthesis pathway in *S. clavuligerus* (Paradkar et al. 1998).

The predicted amino acid sequences of *orf10* (also named *cyp*) and *orf11* (also named *fd*) showed close similarities to a cytochrome P450 and a ferredoxin, respectively (Jensen

et al. 2000; Li et al. 2000). The *orf11* gene is located in the same orientation and downstream of the *orf10* gene (Figure 3). Originally the two gene products were presumed to form a cytochrome/ferredoxin enzymic complex. However, mutants that are defective in *orf11* produce reduced amount of clavulanic acid, whereas mutants that are defective in *orf10* are unable to produce clavulanic acid. This suggests that ORF11 cannot be the only electron transport partner for ORF10 (Jensen et al. 2003a). Besides, mutants disrupted in *orf10* showed elevated levels of production of an unrelated antibiotic holomycin, suggesting a possible complex interregulation between these two antibiotic gene clusters (de La Fuente et al. 2002). To demonstrate the exact role of Orf10 and Orf11 in the clauvlanic acid biosynthetic pathway in *S. clavuligerus*, further experiments are required.

The predicted protein of the *orf12* gene in the clavulanic acid gene cluster showed similarity to two uncharacterized proteins of *Mycobacterium tuberculosis*, to a secreted protein of *Mycobacterium leprae*, and to β-lactamases from *Deinococcus radiodurans*, *Providencia stuartii*, *Streptomyces cacaoi*, *Streptococcus pyogenes* and *Pseudomonas aeruginosa*. Although the protein includes a highly conserved SDN motif of β-lactamases, other residues critical to the catalytic activity of β-lactamases (STFK, EPELN and KGT) are absent. Disruption of *orf12* resulted in mutants with a complete loss of clavulanic acid production ability. However, the role of Orf12 in the biosynthesis of clavulanic acid in *S. clavuligerus* is unknown (Li et al. 2000; Mellado et al. 2002; Jensen et al. 2003a).

The deduced protein of *orf13* presents significant similarity to an export pump for

cysteine and other metabolites of the cysteine pathway, and to an integral membrane protein from *S. coelicolor*. Mutants that are defective in *orf13* produce only about 5% of the wildtype levels of clavulanic acid and *5S* clavams. Thus, Orf13 may act as an export system involved in the excretion of clavam metabolites (Mellado et al. 2002; Jensen et al. 2003a).

The deduced protein of *orf14* shows highest similarity to an acetyltransferase from *Pseudomonas syringae*, a hypothetical protein from *Deinococcus radiodurans*, a hypothetical protein from *S. coelicolor*, and a phosphinothricin acetyltransferase from *D. radiodurans*. All these proteins contain conserved domains characteristic of acetyltransferase. In one study, disruption of *orf14* caused only a partial loss of production (about 33% of wildtype levels, Mellado et al. 2002), whereas in a separate study, *orf14* mutants were almost completely blocked (>99%) in the production of clavulanic acid (Jensen et al. 2003a). This discrepancy cannot easily be explained by differences in disruption methodology. If any, the disruption in the second study should cause a less severe mutation. The role of Orf14 in the biosynthesis of clavulanic acid is unclear.

Disruption of either *orf15* or *orf16* in the clavulanic acid gene cluster in *S. clavuligerus* caused complete loss of clavulanic acid production but appearance of a new metabolite, N-acetylglycylclavaminic acid (Jensen et al. 2003a). The putative ORF15 protein is highly similar (about 50% identity) to the protein sequence of ORF7 in the clavulanic acid gene cluster. However, *orf7* and *orf15* cannot be considered paralogues in

the same way that genes involved in the early steps of clavulanic acid biosynthesis are, mainly due to the following two reasons: Firstly, mutation of either *orf7* or *orf15* causes complete loss of clavulanic acid production, and thus neither of the gene products can compensate for a mutation in the other. Secondly, *orf7* mutant does not make NAG-clavam whereas the *orf15* mutant does. Both lines of evidence imply that they are not paralogue genes. The putative protein encoded by *orf16* shows limited similarities in the protein databases. These include a hypothetical protein D from *Streptomyces hygroscopicus* and a hypothetical protein SF1447 from *Xylella fastidiosa*. The polypeptide encoded by ORF16 contains a conserved hexapeptide ^{375}LPRTGE^{380}, which has been proposed to be responsible for a post-translational modification necessary for the proper anchoring of the proteins to the cell wall (Mellado et al. 2002). Like the *orf15* mutant, clavulanic acid production was lost and NAG-clavam was produced in *orf16* mutants. However, although mutational analysis showed that *orf15* and *orf16* are involved in the clavulanic acid production, their roles are unclear.

The putative protein encoded by *orf17* shows similarity to Y4RH protein (similar to biotin carboxylases) from *Rhizobium* sp., to a carboxylase from *S. coelicolor* and to a carboxylase and an argininosuccinate lyase from *Mesorhizobium loti* (Mellado et al. 2002). The *orf*16 and *orf17* genes were predicted to be translationally coupled, suggesting they are related. Mutants that are defective in *orf17* are unable to produce detectable clavulanic acid (Jensen et al. 2003a), but again, the role of Orf17 in the biosynthesis of clavulanic acid is unclear.

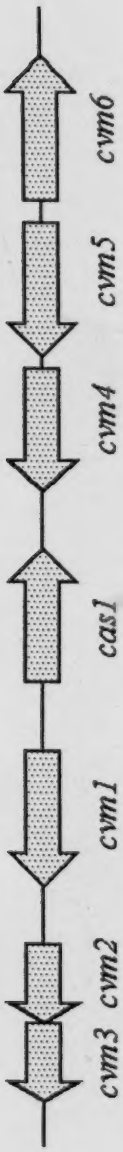
The putative proteins encoded by both *orf18* and *orf19* show strong similarities to penicillin-binding proteins (PBP). Orf18 resembles a type A PBP from different bacteria, including *S. coelicolor*, *S. griseus* and *M. tuberculosis*. Orf19 resembles a type 2 PBP from *S. coelicolor*, *Pasteurella multocida*, *Haemophilus influenzae* and *Rickettsia prowazekii* (Mellado et al. 2002). Disruption of *orf19* showed no apparent effect on viability or metabolite production (Jensen et al. 2003a), while disruption of *orf18* was only possible when a second copy of the gene was provided on a plasmid. This suggests that *orf18* might be an essential gene for the growth of *S. clavuligerus* (Jensen et al. 2003a). A question that arises here is: Are *orfs18 & 19* still part of the clavulanic acid biosynthetic gene cluster or they are associated with primary metabolism as PBPs normally are in other species? Although disruption of *orf19* gene had no marked effects on metabolite production, a possible role for ORF19 in the biosynthesis of clavulanic acid could not be ruled out. Unlike the *orf18* and *orf19* genes which are located side by side at the boundary of the clavulanic acid gene cluster, the genes encoding PBPs in *S. coelicolor* are quite widely distributed throughout the chromosome. In particular, the two PBPs from *S. coelicolor* which show greatest similarity to *orfs18 & 19* (72-76% identity at the amino acid level) are separated by more than 1.5 Mb on the *S. coelicolor* chromosome (Jensen et al. 2003a). Assuming that *orfs18 & 19* are still part of the biosynthetic gene cluster, then the lethal effect of disruption of *orf18* may suggest that the protein encoded by *orf18* might be multi-functional and involved in the biosynthesis of both primary and secondary metabolites. The role of *orfs18 & 19* are unclear and additional genes will have to be

analyzed before it can be concluded that the end of the clavulanic acid gene cluster has been reached. At present, the clavulanic acid gene cluster in *S. clavuligerus* has a total size of about 25 kb including *orf18* and *orf19*.

Although a basic outline for the biosynthesis of clavulanic acid exists, and it is known that clavulanic acid and *5S* clavams share a common biosynthetic pathway up to and including the clavaminic acid step, the details of *5S* clavam biosynthesis beyond the level of clavaminic acid are essentially unknown. However, a correlation was observed between the expression of the *cas1* gene and the production of *5S* clavam metabolites. When the wildtype *S. clavuligerus* is cultivated in SA medium, it efficiently produced clavulanic acid but was unable to produce detectable levels of any of the *5S* clavams. Northern hybridization analysis showed that the *cas1* gene was not expressed in SA medium (Paradkar and Jensen 1995). This correlation might be explained if *cas1* and the other genes for *5S* clavam biosynthesis are coordinately downregulated when *S. clavuligerus* is grown in SA medium (Mosher et al. 1999). With the assumption that *cas1* and *5S* clavam biosynthetic genes are coordinately regulated and thus might be clustered together on the chromosome, genetic regions flanking *cas1* were alanyzed. This revealed the presence of five complete open reading frames *cvm1-5* and one incomplete open reading frame *cvm6* (Figure 5). None of these genes was similar to the open reading frames immediately upstream or downstream of *cas2* in the clavulanic acid gene cluster.

The putative protein encoded by *cvm1* showed a good similarity (>40% identity) to the derived amino acid sequence of a number of putative aldo and keto reductases and

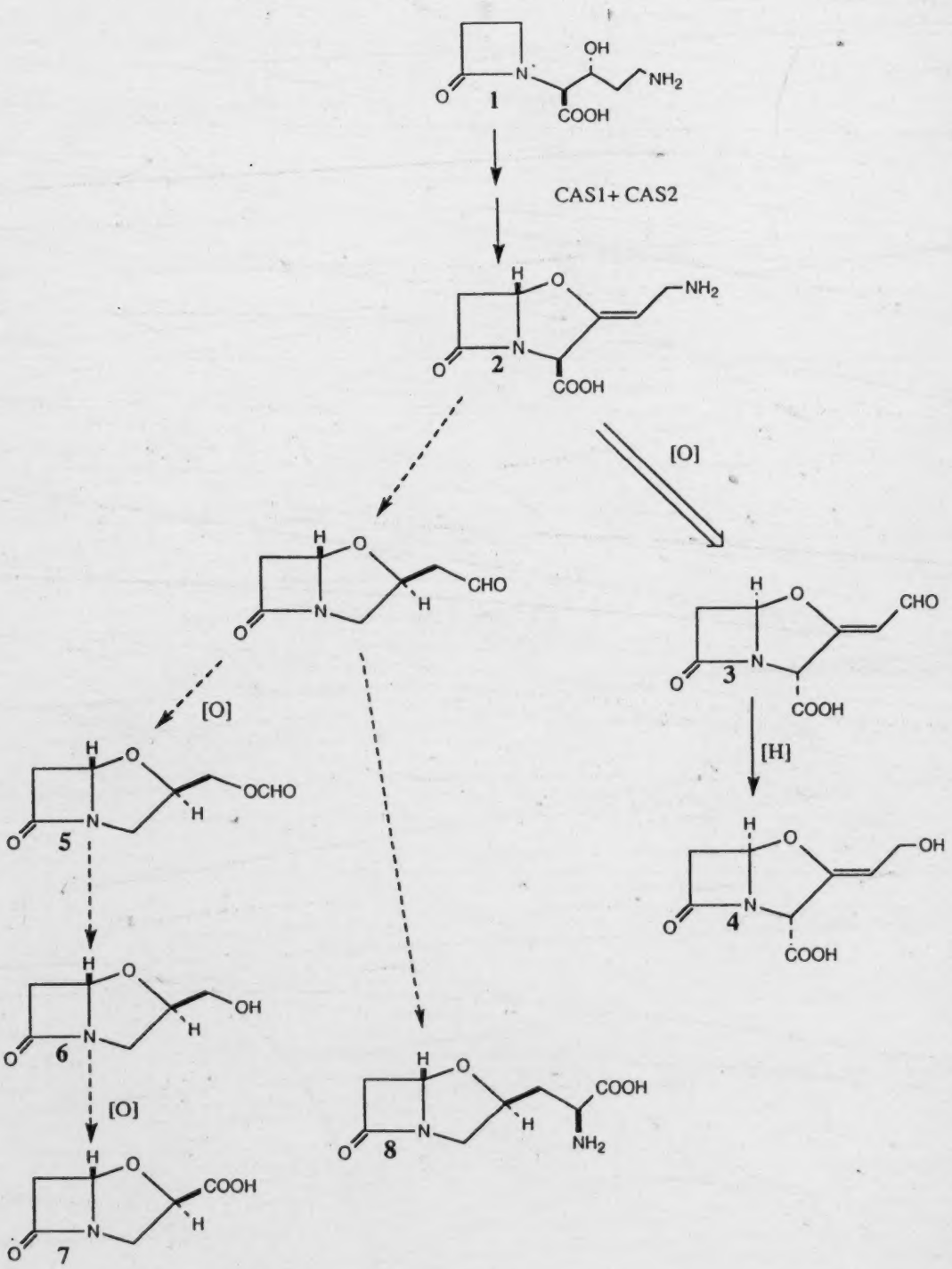
Figure 5. Arrangement of the *5S* clavam biosynthetic gene cluster in *S. clavuligerus* (Mosher et al. 1999). The line in the middle represents the *S. clavuligerus* chromosome DNA. The arrows on the line represent the open reading frames and the direction of their transcription.


cvm3
cvm2
cvm1
cas1
cvm4
cvm5
cvm6

dehydrogenases. The deduced proteins of *cvm2* and *cvm3* do not show any significant similarity to the protein sequences in the database. The putative protein encoded by *cvm4* showed a low degree of similarity to the *cefG*-encoded deacetylcephalosporin C acetyltransferase from *Cephalosporium acremonium*. The putative protein encoded by *cvm5* showed low similarity to the amino-terminal regions of various bacterial luciferases such as that encoded by the *luxA* gene of *Photobacterium leiognathi*. The deduced sequence of *cvm6* showed good similarity to a number of class III pyridoxal phosphate (PLP)-dependent aminotransferases (Mosher et al. 1999).

Mutants that are defective in *cvm1* and mutants that are defective in both *cvm4* and *cvm5* still produced clavulanic acid and cephamycin C but lost the ability to synthesize the *5S* clavam metabolites clavam-2-carboxylate, 2-hydroxymethyl-clavam and 2-alanylclavam. These results agree with the hypothesis that *cas1* is clustered with genes essential and specific for clavam metabolite biosynthesis (Mosher et al. 1999). However, the exact role of the proposed products encoded by these genes is uncertain, primarily because little is known about the biochemistry of the pathway leading from clavaminic acid to the various clavam metabolites produced by *S. clavuligerus*. However, the derived amino acid sequence of *cvm6* which possessed significant similarity to a variety of PLP-dependent aminotransferases might act as the PLP-dependent enzyme(s), as proposed by Egan et al. in the tentative biosynthetic pathway (Egan et al. 1997, Figure 6), to decarboxylate and deaminate clavaminic acid to form various clavam metabolites (Mosher et al. 1999).

Figure 6. Proposed pathway for the biosynthesis of clavulanic acid and the *5S* clavams beginning at proclavaminic acid (Mosher et al. 1999). 1, proclavaminic acid; 2, clavaminic acid; 3, clavulanate-9-aldehyde; 4, clavulanic acid; 5, 2-formyloxymethylclavam; 6, 2-hydroxymethylclavam; 7, clavam-2-carboxylic acid; 8, 2-alanylclavam.


1
OH
NH2
COOH
CAS1+ CAS2
2
NH2
COOH
[O]
CHO
3
COOH
[H]
4
OH
COOH
[O]
OCHO
5
[O]
6
OH
[O]
7
COOH
8
COOH
NH2

Disruption of structural genes involved in the biosynthesis of particular *5S* clavams should result in selective loss of these *5S* clavam metabolites, however, so far, no such genes have been found. Thus it is necessary to search for the biosynthetic genes associated with the formation of specific *5S* clavam metabolites. Since antibiotic biosynthetic genes are usually clustered in *Streptomyces*, the further flanking regions of *cas1* and the flanking regions of the paralogue gene cluster are logical genetic regions to search for those genes, and the potential involvement of those genes can be assessed by gene disruption.

One of the traditional ways to prepare gene disruption mutants in *S. clavuligerus* is, in brief, to disrupt the cloned genes by insertion of an antibiotic resistance marker into a restriction site within the gene with at least 500 bp of flanking DNA on each side. Typical antibiotic resistance markers that are used to mark the disruption in *S. clavuligerus* are Apra (apramycin), thiostrepton and neomycin resistance genes. To deliver the disrupted gene fragments, a pIJ101-based *Streptomyces* plasmid vector is used in our laboratory. Such plasmids are segregationally unstable in *S. clavuligerus* in the absence of antibiotic selection. After transforming the plasmids into wildtype *S. clavuligerus*, transformants are allowed to sporulate in the absence of antibiotic, and thus result in the loss of free-living plamids. Disruption mutants arising from double crossover are screened by antibiotic resistance, and confirmed by Southern analysis by using DNA fragments of the antibiotic resistance gene as a probe. This gene disruption process is laborious and time consuming mainly because of the requirement of the cloning the gene to be disrupted and subsequent

sequencing analysis by primer walking to identify a suitable restriction site within the gene. In addition, the slow growth rate of *Streptomyces* spp. and the cumbersome transformation procedures used with *Streptomyces* spp. usually take at least three months to prepare disruption mutants in a single gene.

Compared with the traditional gene disruption methods, creation of gene disruption by transposon insertion (transposon mutagenesis), showed marked advantages in many aspects. This includes the facile process involved, no requirement for DNA sequences and gene function information, large number of transposon-containing mutants generated after one simple transposition reaction, capability of selection and transfer of mutant clones by using the antibiotic resistance marker of the transposon, and availability of primer-binding sites for DNA sequence analysis so that large DNA fragments can be sequenced without primer walking. Therefore, to simplify the generation of disruption mutants, we plan to explore the approach of *in vitro* transposon mutagenesis in our research.

Transposons, or transposable genetic elements, or "jumping genes", are DNA fragments capable of moving from one place to another in the genome independent of extensive DNA homology. The first observations regarding the existence of transposable elements were made about 70 years ago. It was found that certain DNA fragments could lead to an increase in spontaneous mutation frequencies of other regions within the cell. Those DNA fragments were originally described as “Mutation Genes”. In the late 1940s, Barbara McClintock described “Controlling Elements” in maize. Those elements were

shown to be responsible for the diversity of corn seed colors and a degree of other genetic instability in plants. During the 1960s, it became apparent that bacterial genetic systems, like their eukaryotic counterparts, were also subject to genetic instability. In 1974, Hedges and Jacob observed the first bacterial transposition phenomenon that led to the transfer of Amp resistance from the RP4 plasmid to other replicons within the cell (Hedges and Jacob 1974).

Transposable elements are now known to be found throughout nature, being common in bacteria and plants. Most transposable elements are resident on plasmids or incorporated into bacteriophage DNA, facilitating their dispersal. Some are found upon the bacterial chromosome, but in any case there is usually only one copy present on the same replicon. This is controlled by a mechanism named "transposon immunity", whereby the presence of one transposon on a replicon makes the replicon immune to the integration of a second copy of the same element. This can be interpreted as transposons tend not to integrate into themselves since that would be suicidal. Also, two proximal copies of a transposon on the same DNA could then cause deletion of the sequences between them by homologous recombination.

With several exceptions such as IS*91* and IS*110*, the majority of transposable elements exhibit terminal inverted repeats (IRs) at their ends. IRs generally contain two functional domains: domain A includes two or three terminal DNA base pairs involved in cleavage and strand transfer reactions leading to transposition of the element; domain B provides the transposase binding site (Ichikawa et al. 1990). Short direct repeats (DRs)

flanking the transposon are the hallmarks for the occurrence of transposition events. Derived from target DNA, the length of DRs is usually between 2 and 14 bp. Their presence can be explained by the attack of the end of the transposon to each target DNA strand in a staggered way.

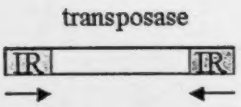
Transposase (Tnp), encoded by the *tnp* gene, is resposonsible for catalyzing transposition reactions. There is a general pattern for the domain structure of most transposases: the N-terminal region is the sequence-specific DNA binding domain, while the catalytic domains are often present at the C-terminal end (Haren et al. 1999).

There are three general types of bacterial transposable elements (Figure 7). The simplest ones are insertion sequences or IS elements. Discovered during early investigation of gene expression in *E. coli* and bacteriophage lambda through their repetitive presence within mutant genes, IS elements share several properties including: frequently less than 1 kb in length, contain terminal IRs, carry no detectable markers such as antibiotic resistance genes other than those required for transposition reaction or regulation within these elements. However, their lack of identifiable genetic markers makes them the most difficult transposable elements to be studied, and they are generally investigated indirectly through their ability to cause mutations.

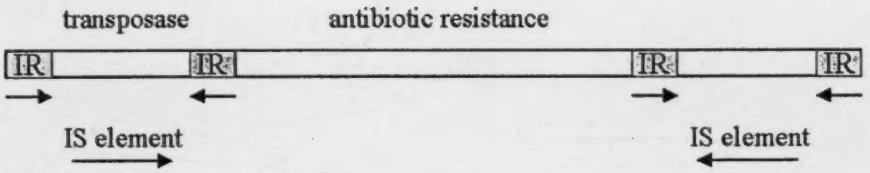
When two IS elements of the same type occur on the same replicon, they form a composite (or compound) transposon. Unlike IS elements, composite transposons encode auxiliary genes, such as drug resistance genes, which makes them easier to be isolated and transferred. The most common arrangements of IS elements in composite transposons

Figure 7. Structures of bacterial transposable elements. IR: inverted repeats.

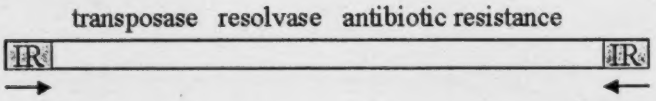
IS elements



Composite transposons



Complex transposons

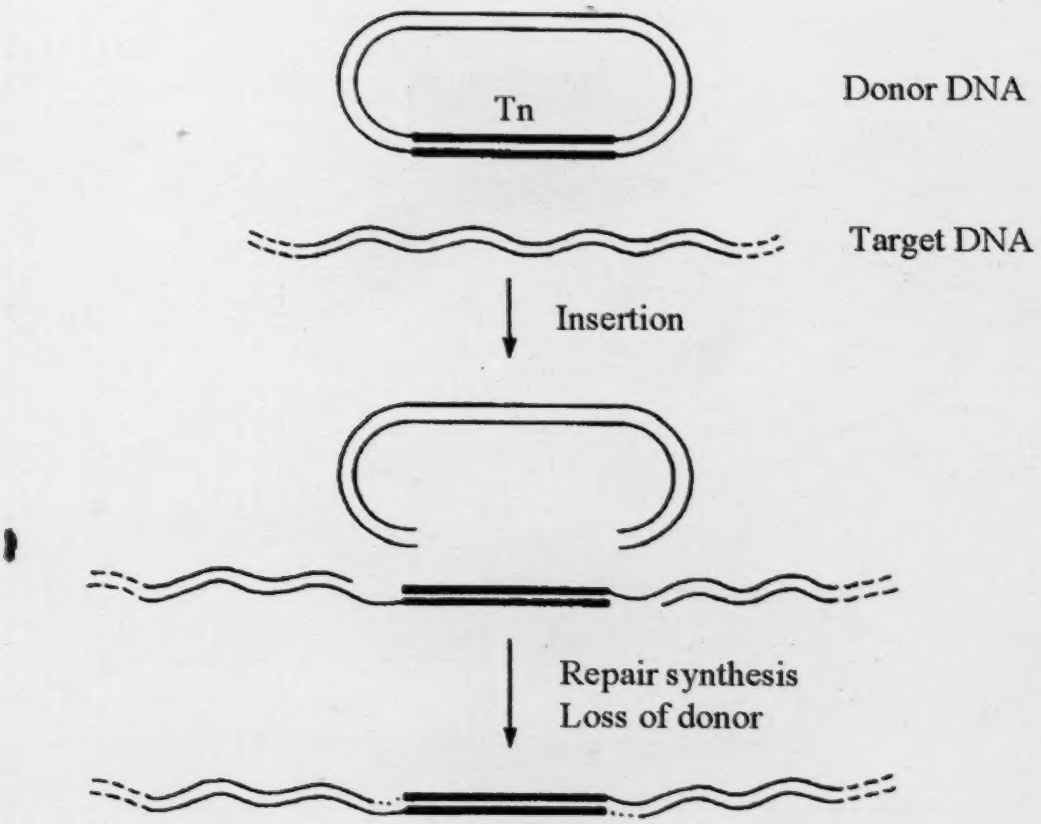


are in IRs such as Tn*5* and Tn*903*, while some are present in DRs such as Tn*9* (Snyder and Champness 1997). Transposition function in composite transposons is supplied by the IS elements. By encoding transposase genes and acting as the recognition sites for transposase, IS elements excise and transpose the composite transposon as a whole unit. Mutations occurring in one of the IS elements in composite transposons are commonly found. In this case, only one IS element will provide the transposition function.

The third major category of bacterial transposable elements is complex transposons, which contain two IRs rather than insertion sequences at their extremities. Tn*1* or Tn*3*, both of which contain two 38 bp long identical IRs, can exemplify this type of transposons. Transposition function of complex transposons is generally supplied by their central sequences, which carry all the genes required for transposition, including *tnp*, the resolvase encoding gene *tnpR*, and the genes encoding detectable markers (Berg et al. 1989).

Although the basic transposition features seem to be shared by all transposable elements, the detailed mechanisms, frequencies, and distributions of transposition products are element-specific. All bacterial transposable elements seem to use either conservative transposition or replicative replication for their replication. Conservative (or Cut and Paste) transposition (Figure 8), exemplified by Tn*5* and Tn*10*, involves physical removal of transposable elements from one replicon and insertion into another within the same cell. During a conservative transposition reaction, the transposase cleaves both strands at each vector–transposon junction, and thus the transposon is excised from the

Figure 8. Mechanism of conservative transposition (modified from Snyder and Champness 1997).

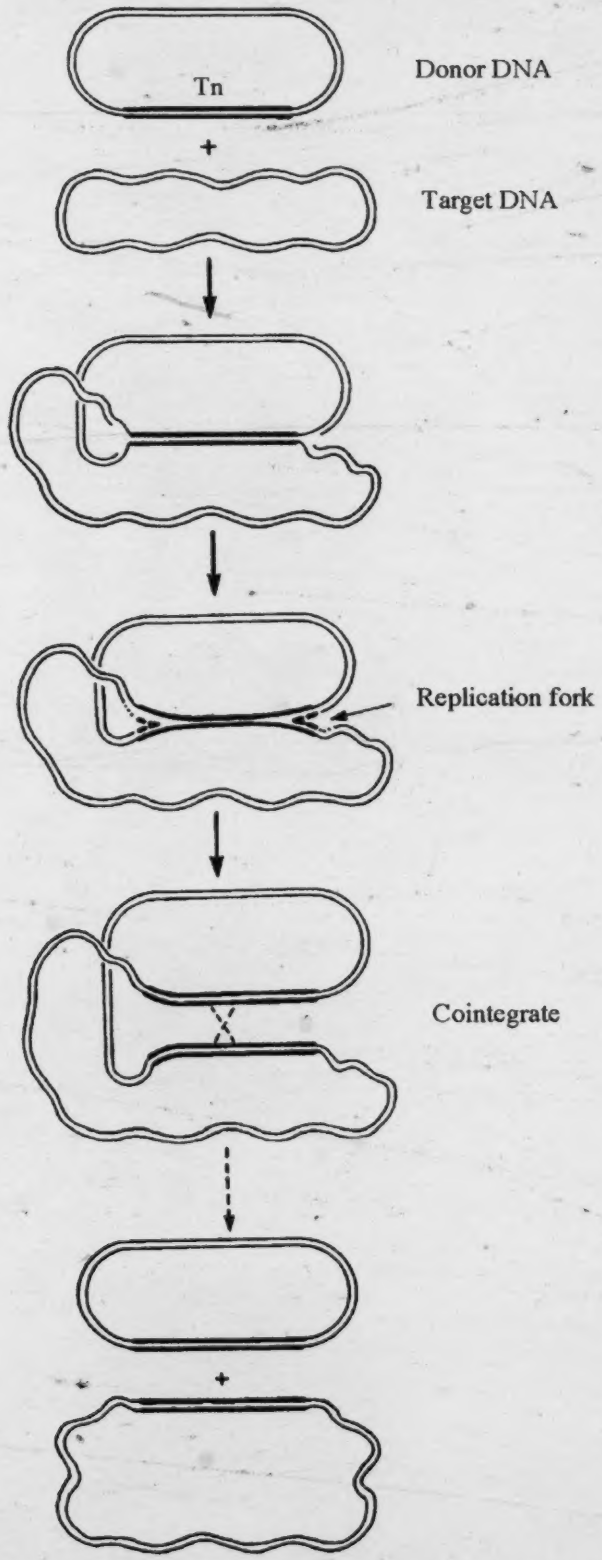

Tn
Donor DNA
Target DNA
Insertion
Repair synthesis
Loss of donor

original site leaving a double-strand break behind, which is irreparably damaged as a result. At the same time, staggered cuts are introduced into the target site. After ligating with the released transposon, the single stranded DNA gaps left by the staggered cuts are filled by the DNA repairing system of the cell. As a result of the conservative transposition reaction, the transposable element leaves the donor molecule and appears in the recipient molecule, and is flanked by the direct repeats of duplicated target sequence. Thus the copy number of the transposon is conserved during transposition.

Replicative transposition, exemplified by Tn*3*, is a mechanism whereby the element replicates itself from place to place (Figure 9). During a replicative transposition reaction, transposase cleaves only one strand at the vector-transposon junction and staggered cuts are introduced into the target site. The transposon donor and recipient molecules now fuse and generate a cointegrate structure that integrates the entire donor molecule into the recipient molecule. Some transposable elements, such as IS*6100* (Smith and Dyson 1995) and Tn*1792* (Herron et al. 1999), form cointegrates as an end product of transposition. For others, however, cointegrate formation can then be resolved through the action of host-encoded or transposon-encoded resolvase, generating the genetically unchanged donor molecule and a target molecule with a copy of the transposable element inserted. Since the replicative transposition mechanism does not destroy donor DNA, it differs from conservative transposition mechanism in the replication of transposons in the cell.

Transposable elements differ greatly in their target specificity. None of them can completely randomly insert into target DNA. Some transposable elements display a

Figure 9. Mechanism of replicative transposition (modified from Snyder and Champness 1997).


Tn
Donor DNA
+
Target DNA
Replication fork
Cointegrate
+

marked preference to insert into specific consensus sequences. For example, IS*50* transposes preferentially insert into 5'-GNTYWRANC, where N=all 4 bases, Y=T or C, W=A or T, and R=A or G. (Goryshin et al. 1998). Tn*10* often inserts into hotspots containing the symmetrical consensus sequence 5'-GCTNAGC, where N=all 4 bases. However, this consensus is not sufficient to determine its insertion specificity. Mutational analysis revealed that the 6-9 bp flanking sequences of each end of the consensus sequence also contribute to Tn*10*'s target specificity (Bender and Kleckner 1992). For some transposable elements, specific target sequences seem less important than local properties of the DNA such as DNA bending (IS*10*, Pribil and Haniford 2000), DNA replication (Tn*7*, Wolkow et al. 1996) or transcription (Tn*7*, DeBoy and Craig 2000) for their target selection. Although Tn*5* is among the most random of the known transposable elements, its target sites also show hotspots that have little resemblance to one another. Mutational analysis revealed several features that influence the occurence of Tn*5* insertion, including terminal GC pairs of the target sites, and some regional features of DNA such as the transcriptional activity through the site and negative DNA supercoiling (Berg 1989).

By virtue of their ability to move as discrete DNA segments to new locations and disrupt their target DNA, natural and engineered transposable elements are used as molecular genetic tools by *in vivo* or *in vitro* transposition reactions—transposon mutagenesis. As an important tool for genetic and molecular analysis, transposon mutagenesis has several advantages over other genetic or molecular approaches: (1).

Differing from other gene disruption methods such as gene replacement, target DNA selection of transposon mutagenesis is mainly based on the characteristics of transposon itself, therefore knowledge of DNA sequences or gene function is not required. (2). The drug resistance gene derived from a composite or complex transposon can be used for both transposon insertion selection and mutation transfer. (3). Once the transposon mutant with the phenotype of interest is isolated, known restriction sites and sequences of transposons can be used as physical or genetic landmarks to identify transposon-flanking sequences. (4). Transposon based expression systems can facilitate gene expression analysis. For instance, a Lux-encoding *Streptomyces* transposon Tn*5353* that employs the *Vibrio harveyi* luciferase-encoding *luxAB* cassette as a reporter of transcription was used for screening of *S. coelicolor* promoters that exhibit developmental phenotypes or that are induced by various environmental stimuli (Sohaskey et al. 1992). (5). Transposons can also serve as mobile promoters to control adjacent genes. For instance, a Tn*5* derivative transposon, Tn*5lacIq*/*Ptrc*, containing an outward facing, *lacI*-regulated promoter (P_{trc}) was constructed and was used to analyze expression of *Pseudomonas* genes (de Lorenzo et al. 1993).

To date, various *in vivo* and *in vitro* transposition systems have been developed and applied. The *in vivo* transposition systems have been used for transposon mutagenesis by introducing transposable elements and the *tnp* gene into recipient cells through a plasmid or phage suicide vector. Such suicide vectors usually have one or more defects so that they cannot replicate in a particular host or under a particular set of conditions. For

instance, some have a narrow host range for replication; some have mutations in the vector or mutations in the recipient cell that makes the vector replication deficient. They can be propagated in a permissive host or under a permissive condition before introducing into the transposition recipient cell or under non-permissive condition. However, the only way for a stably drug-resistant cell to exist is through the association of the transposon with a replicon in the recipient cells by transposition.

Plasmid delivery systems take advantage of a transposon insertion on a plasmid that is unable to replicate in the recipient cell. Theoretically, any plasmid with a conditional lethal mutation, either nonsense or temperature sensitive, or narrow host range can be used as a suicide vector. For instance, a temperature sensitive plasmid vector pSIT151 was used to deliver the mini transposon Tn*1792* from *E.coli* to *S. lividans* and *S. avermitilis* to create *in vivo* transposon mutagenesis (Herron et al. 1999; Pitman et al. 2002). Phage delivery systems take advantage of a transposon insertion on a phage that is unable to lysogenize the recipient cell. For instance, derivatives of phage λ that cannot insert into the chromosome due to the deletion of their attachment sites, and those containing conditional defects because of nonsense mutations in replication genes, are commonly used as transposon vectors in *E.coli*.

All *in vivo* transposition systems have a number of technical limitations: typically they are performed species specifically; the transposable element must be introduced into the host through a suicide vector; the transposase must be expressed in the target host; but the transposase usually is expressed in subsequent generations, resulting in potential

genetic instability. The inherent difficulty of use of *in vivo* transposition systems has been reduced by the recent development of *in vitro* transposition systems. These systems involve a host independent, test-tube-based reaction system including recipient DNA, a transposon and its corresponding transposase. Most *in vitro* transposition systems are used to mutagenize DNA fragments cloned in plasmid or cosmid vectors. For instance, a Tn*5* based *in vitro* transposition system has been used to generate gene mutations in *S. coelicolor* cosmid library DNAs (Gehring et al. 2000).

In vitro and *in vivo* transposition have been linked by the use of Transposome. It is a complex of transposase and transposon, generated during the early steps of *in vitro* transposition when the transposase attaches to the transposon in the absence of recipient DNA or cations to activate transposition. To be used for transposon mutagenesis, Transposomes have to be introduced into recipient cells, and then transpose into the host genome. Thus the technical difficulties of this method lie in the transformation of Transposomes into recipient cells. A currently well-developed example of Transposome is based on Tn*5*. By electroporation of Tn*5*-based Transposome into host organisms, Tn*5* insertions have occurred on the chromosomes of *E. coli*, *Salmonella typhimurium*, and *Proteus vulgaris* (Goryshin et al. 2000).

To apply transposon mutagenesis in a research project, several factors need to be considered: (1). The choice of transposon should be guided by the goals of the project. For example, if the purpose of the project is to obtain one or a few insertions within a 100 kb region, then insertion hotspots will not be a limitation. However, if an insertion is

needed for every few hundred base pairs of DNA fragments, the randomness of insertion should be the major concern. (2). Careful selection of a resistance marker. Some resistance genes work poorly in certain organisms, and even affect the randomness of transposon insertion. For instance, *Pseudomonas aeruginosa* is sensitive to tetracycline, however, Tn*10* shows high target site specificity when used with tetracycline selection and thus cannot be applied to this system. (3). Organism specificity. Most transposition systems borrow certain host machinery during transposition, and therefore they tend to be limited to hosts with "compatible machinery". (4). Stability of insertions. Although mutations induced by most transposable elements are stable, it is still possible to have part or all of the transposon excised from its insertion site. Alternatively, the transposon may transpose from the original site to a new site. Reversion frequency depends on a number of factors, such as the insertion site, whether it is on an F plasmid, and strain background. (5). Other factors include the potentially questionable transposon mutagenesis of genes in which production of a portion of the gene is sufficient to provide function; and an inability to assess the contribution of a gene that is duplicated or has a functional paralogue.

At the end of the 1980s, transposon mutagenesis was added to the toolbox for *Streptomyces* genetic manipulation with a number of potential applications, including tagging secondary metabolite, sporulation, regulatory and other genes; activating cryptic genes by promoter insertions; up-regulating genes involved in the biosynthesis in

secondary metabolite production by promoter insertions; and providing portable regions of homology for recombination between different species or genera (Baltz et al. 1997).

The first *Streptomyces* transposon, IS*110*, was found in 1985 from *Streptomyces coelicolor* A3(2) (Chater et al. 1985). Two years later, Tn*4556*, was found from *Streptomyces fradiae* when it jumped onto the plasmid prophage, SF1 (Chung 1987). Since then, several other transposable elements have been discovered in *Streptomyces* spp., but none of them contain a selective marker and thus can not be applied for transposon mutagenesis. For easier manipulation of these transposable elements, antibiotic resistance genes have been added *in vitro*. For instance, genetically engineered Tn*4556*, constructed by inserting a viomycin resistance gene and a lincomycin resistance gene (Chung 1987), was used to mutagenize a cloned *whiE* locus in *S. coelicolor* (Davis and Chater 1990). Its derivative, Tn*4560*, was constructed by inserting a viomycin phosphotransferase gene (*vhp*). Tn*4560* was used to tag avermectin biosynthetic genes in *S. avermitilis*, which were then cloned and expressed in other strains (Ikeda et al. 1993). In addition, Tn*4560* was used to analyze genes involved in nikomycin biosynthesis in *Streptomyces tendae* (Engel and Wright 1993, Engel and Lax 1997, Engel et al. 2001).

Several derivatives of IS*493* were designed for transposon mutagenesis. Tn*5096* was constructed by inserting an Amp resistance gene into IS*493*. Tn*5096* transposes relatively randomly in several *Streptomyces* species and has contributed to gene isolation and functional research (Gagnat et al. 1999). Tn*5099* is a promoter probe transposon, constructed by inserting a cassette containing a promoterless *xylE* gene encoding catechol

dioxygenase that gives yellow colonies in the presence of catechol. The daptomycin biosynthetic gene cluster of *S. roseosporus* was analyzed by Tn*5099* mutagenesis (Baltz et al. 1997).

Although several indigenous transposable elements have been modified and applied in gene disruption, the applications of transposon mutagenesis in *Streptomyces* spp. are still quite limited. This limited use is mainly due to the fact that most indigenous *Streptomyces* transposable elements exhibit high target site specificity. Some transposable elements within the genus can transpose relatively randomly in certain species, but they usually function suboptimally or not at all in other species. This results in great difficulties in the broad application of transposon mutagenesis within the *Streptomyces* genus.

Application of heterologous transposable elements in *Streptomyces* seems even more difficult due to host incompatibility. However, in the past seven years, two transposition systems involving heterologous transposoable elements have been developed and successfully applied. This includes IS*6100* from *Mycobacterium fortuitum* and its derivative Tn*1792*, and Tn*5* from *E.coli*, providing valuable information for recruiting heterologous systems for the application of transposon mutagenesis in *Streptomyces*.

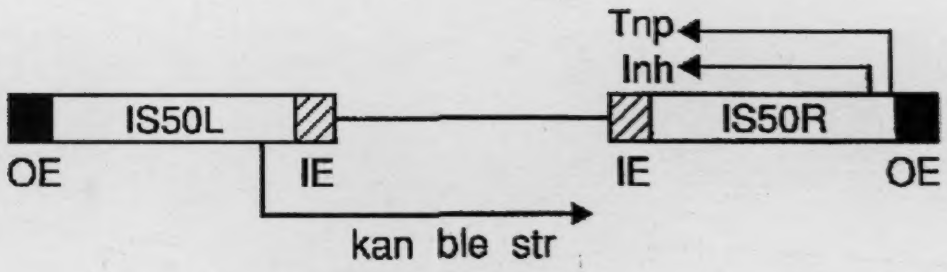
IS*6100* was isolated as a component of the composite transposon Tn*610* from *M. fortuitum*. IS*6100* transposes replicatively and forms cointegrate as an end product of transposition. IS*6100* was used for *in vivo* transposon mutagenesis in *S. lividans* (Smith and Dyson 1995), and *S. avermitilis* (Weaden and Dyson 1998). Its derivative, Tn*1792*

was constructed through introduction of an *aacC1* gene conferring gentamicin resistance between two IS*6100* IRs. Tn*1792* was used for *in vivo* transposon mutagenesis in *S. lividans*, revealing low target site specificity and formed a stable cointegrate structure (Herron et al. 1999). However a high frequency of cointegrate resolution has been observed when Tn*1792* transposes into *S. avermitilis* (Pitman et al. 2002).

E.coli Tn*5* (Figure 10), a 5.7 kb composite transposon, shows low target site specificity and is applied in many Gram-negative bacteria, and in an *in vitro* transposition reaction with *S. coelicolor* cosmid library DNAs. In this experiment, insertions were identified at widely scattered locations and revealed several novel mutant phenotypes, such as altered pigment production, enhanced formation of aerial hyphae, and a block in spore formation (Gehring et al. 2000). This represents the first example of applying *in vitro* transposition in *Streptomyces*.

Tn*5* contains two nearly identical inverted versions of IS*50* bracketing three antibiotic resistance genes encoding resistance to Km (kanamycin), bleomycin and streptomycin (Figure 10). The right IS*50* (IS*50*R) encodes the Tnp and the transposase inhibitor (Inh). The left IS*50* (IS*50*L) encodes C-terminal truncated, inactive versions of Tnp and Inh and contains the promotors for the antibiotic resistance genes. The IS*50* elements are defined by two 19 bp sequences termed the outside end (OE) and inside end (IE), which are critical Tnp binding sites. The OE and IE differ at seven positions. The internal DNA encompassed by the 19 bp end sequences can be any DNA sequence. It need only be long enough to allow the two ends to come in close proximity so that Tnp

Figure 10. Tn*5* structure (Reznikoff 2003). OE: outside end; IE: inside end; kan: kanamycin-resistance gene; ble: bleomycin-resistance gene; str: streptomycin-resistance gene.


Tnp
Inh
IS50L
IS50R
OE
IE
IE
OE
kan ble str

bound to the end sequences can form a synaptic complex. Tn*5* transposes with high frequency via a conservative transposition mechanism. After transposition, it generates a 9-bp DR derived from the target sequence at the site of insertion (Reznikoff 2003).

The wildtype Tn*5* Tnp is a very inactive protein. Purified wildtype Tnp shows no evidence of transposition activity *in vitro*, and *in vivo* the transposition frequency is on the order of 1 event in 10^5 cells, even though abundant protein is present. The biochemical basis for this inactivity has been studied through the isolation of hyperactive mutations. These mutations include introduction of a α-helix to facilitate the conformational change of the protein; enhancement of the DNA-Tnp interaction; exposure of the active site of Tnp, etc (Reznikoff, 2003). Construction of hyperactive Tn*5* transposase makes Tn*5* transpose highly efficiently.

Limited by the availability of transposable elements that can both function and show low target site specificity in *Streptomyces* spp., transposon mutagenesis has only been applied to a few species of *Streptomyces* excluding *S. clavuligerus*. However, genetically engineered transposition systems, e.g. Tn*5*, provide promising prospects for broader application of this approach in the genus. Compared with the traditional laborious and time-consuming gene disruption methods used in our lab, transposon mutagenesis shows marked advantages in many aspects as described above, especially the large number of disruption mutants generated all at once, and the primer-binding sites provided by the transposon that greatly facilitate the DNA sequencing process. To analyze *ceaS1*-side flanking region of the paralogue gene cluster, and facilitate the generation of a large

number of gene mutants, the Tn*5* *in vitro* transposon mutagenesis system was used on cosmid 6G9, which carries the cloned paralogue gene cluster and its flanking regions from *S. clavuligerus,* to create gene disruptions. By introduction of transposon-containing cosmids into wildtype *S. clavuligerus*, *S. clavuligerus* mutant strains were constructed, and the involvement of the disrupted genes in clavam biosynthesis was assessed by bioassay and HPLC.

II. MATERIALS AND METHODS

II.1 Materials

Restriction endonucleases were purchased from Boehringer Mannheim Biochemicals (Laval, QC), New England Biolabs (Mississauga, ON), or Promega Corp. (Madison, WI). *E. coli* DNA polymerase I was purchased from Roche Diagnostics Corp. (Indianapolis, IN). DNA polymerase from *Thermus aquaticus* (*Taq*) and DNAse I were purchased from Boehringer. Transposase was purchased from Epicentre (Madison, WI). Lysozyme was purchased from Sigma Chemical. CO. (St. Louis, MO). All enzymes were used according to specifications advised by the manufacturers.

Lambda DNA was purchased from Amersham Pharmacia Biotech. Inc. (Piscataway, NJ). Ultrapure DNA-grade agarose was obtained from ICN Biochemicals Inc. (Aurora, OH). Deoxyribonucleoside triphosphates (dNTPs) were purchased from Amersham. Antibiotics were purchased from Sigma. Radioactively labeled [α-^{32}P]dCTP was obtained from Amersham. Clavam-2-carboxylate, clavaminic acid and clavulanic acid standards were gifts from Glaxo Smithkline, United Kingdom.

ISP4, Trypticase Soy Broth (TSB), Bacto Yeast Extract and Bacto Tryptone were purchased from Becton, Dickinson and Company (Sparks, MD). Soybean Flour was purchased from Sigma.

Buffers and other chemicals used were purchased from ICN, Sigma, Life Technologies (Gaithersburg, MD), BDH (Toronto, ON) or EM Science (Gibbstown, NJ).

QIAquick gel extraction kit was used to elute DNA from agarose gels. The kit was

purchased from QIAGEN Inc. (Mississauga, ON).

All the oligonucleotide primers used in this study were designed by the software GeneTool version 1.0 (BioTools, Edmonton, AB). The primers used to sequence the transposon flanking regions and the primers used to prepare probes for determination of the *Eco*RI fragment of interest on 6G9 were ordered from the DNA Synthesis Laboratory at Department of Biological Sciences, University of Alberta. The primers used for checking sequencing errors were ordered from QIAGEN (Alameda, CA).

II.2 Bacterial strains, growth media, culture conditions, plasmids and cosmids

II.2.1 Bacterial strains, growth media and culture conditions

The list of bacterial strains used in this study is given in Table 1.

Both wildtype *S. clavuligerus* 3585 and the *S. clavuligerus* mutant strains constructed in this study were maintained temporarily on ISP4 plates. For long-term maintenance, spores were scraped from plates using a sterile spatula, washed with sterile water, resuspended in 20% (v/v) glycerol and stored as frozen glycerol stocks of spores at -70°C.

For the purpose of genomic DNA extraction from *Streptomyces*, flasks (125 mL) containing 25 mL TSBS (TSB + 1% starch) media were inoculated with 20 µl of frozen glycerol stocks of spores. The cultures were incubated on a 28°C shaker at 250 rpm for 36-40 hr.

For β-Lactam bioassays and HPLC analysis of antibiotic production, seed cultures of both *S. clavuligerus* 3585 and *S. clavuligerus* mutants constructed in this study were started from the frozen glycerol stocks of spores as described for genomic DNA

Table 1. List of bacterial strains used in this study.

Strains	Descriptions and References
E. coli	
TransforMax™ EC100™	Δ*endA1*, Δ*recA1* and highly transformable, used as transformation host for preparation of transposon-containing clones. From Epicentre (Markham, ON).
ET12567 (pUZ8002)	Methylation-deficient and mobilization-proficient (Δ*dam*, Δ*dcm* and Δ*hsdS*), containing a helper plasmid pUZ8002, a RK2 derivative with defective *ori*T which supplies the RK2 transfer function. During cultivation, Km is needed to maintain selection for pUZ8002, and Cm is needed to maintain selection for *dam* mutation (MacNeil 1988; Paranthaman and Dharmalingam 2003). From Dr. Leo C. Vining (Dept. of Biology, Dalhousie University, Halifax).
DH5α	F^-, used as transformation host (Hanahan 1983).
ESS	Cephamycin C sensitive, used as an indicator organism for cephamycin C bioassays (Aharonowitz and Demain 1978). From Dr. Arnold L. Demain (Charles A. DANA Research Institute, Drew University, Madison, NJ).
S. clavuligerus	
3585	Wildtype strain. From Northern Regional Research Laboratories (now the National Center For Agricultural Utilization Research).
A2	Mutant strain with disruption in the middle of *orfB* of the 11.2 kb *Eco*RI fragment of cosmid clone 6G9. This study.
B50	Mutant strain with disruption in the 5' end of *orfB* of the 11.2 kb *Eco*RI fragment of cosmid clone 6G9. This study.
D41	Mutant strain with disruption in *orfC* of the 11.2 kb *Eco*RI fragment of cosmid clone 6G9. This study.
B7	Mutant strain with disruption in the middle of *orfI* of the 11.2 kb *Eco*RI fragment of cosmid clone 6G9. This study.

Table 1. Continued...

Strains	Descriptions and References
A5	Mutant strain with disruption in the 5'-end of *orfJ* of the 11.2 kb *Eco*RI fragment of cosmid clone 6G9. This study.
D10	Mutant strain with disruption in *orfJ* of the 11.2 kb *Eco*RI fragment of cosmid clone 6G9. This study.
B20	Mutant strain with disruption in *orfK* of the 11.2 kb *Eco*RI fragment of cosmid clone 6G9. This study.
B42	Mutant strain with disruption in *orfL* of the 11.2 kb *Eco*RI fragment of cosmid clone 6G9. This study.
IC4	Mutant strain with disruption in *orfA* of the 2.8 kb *Eco*RI fragment of cosmid clone 6G9. This study.
Klebsiella pneumoniae ATCC 29665	β-lactamase producer, used as an indicator organism for clavulanic acid bioassays (Nathisuwan et al. 2001). From ATCC (American Type Culture Collection).
***Bacillus* spp.** ATCC 27860	Clavam sensitive, used as an indicator organism for alanylclavam bioassays (Pruess and Kellett 1983). From ATCC.

extraction from *Streptomyces*. The cultures were incubated on a 28°C shaker at 250 rpm for approximately 40 hr. Seed cultures were used to inoculate 25 mL amounts of Soy medium to 2% (v/v), dispensed in 125 mL flasks. Soy medium contained: soy flour 15 g, soluble starch 47 g, KH_2PO_4 0.1 g, and $FeSO_4 \cdot 7H_2O$ 0.005 g per liter adjusted to pH 6.8. Filter-sterilized $FeSO_4 \cdot 7H_2O$ was added after the medium had been sterilized in an autoclave. The Soy cultures were then incubated on a 28°C shaker at 250 rpm for up to 96 hr. The cultures were harvested at 72-hr and 96-hr time points, and the culture supernatants were used for β-Lactam bioassays and HPLC analysis.

When *Streptomyces* cultures carried cosmids or transposon-containing clones with the apramycin (Apra) resistance gene, Apra was added to a final concentration of 25 μg/mL.

E. coli cells were cultivated on LB broth supplemented with appropriate antibiotics to maintain the plasmid, cosmid, or transposon-containing clones. LB broth contained: tryptone 16 g, yeast extract 10 g, and NaCl g per liter. Frozen glycerol stocks of *E. coli* cells were prepared by harvesting overnight cultures, resuspending the cells in 20% glycerol, and preserving the cells at –20°C for short-term, or -70°C for long-term storage.

To isolate plasmid or cosmid DNA from *E. coli*, one isolated colony was used to inoculate 2 mL of LB broth supplemented with appropriate antibiotics. The cultures were incubated at 37°C for 14-18 hr on a tube roller. To prepare *E. coli* electrocompetent cells, an overnight culture was used to inoculate fresh LB broth to 1% (v/v). The culture was then incubated at 37°C until reaching an OD_{600} of 0.35-0.5.

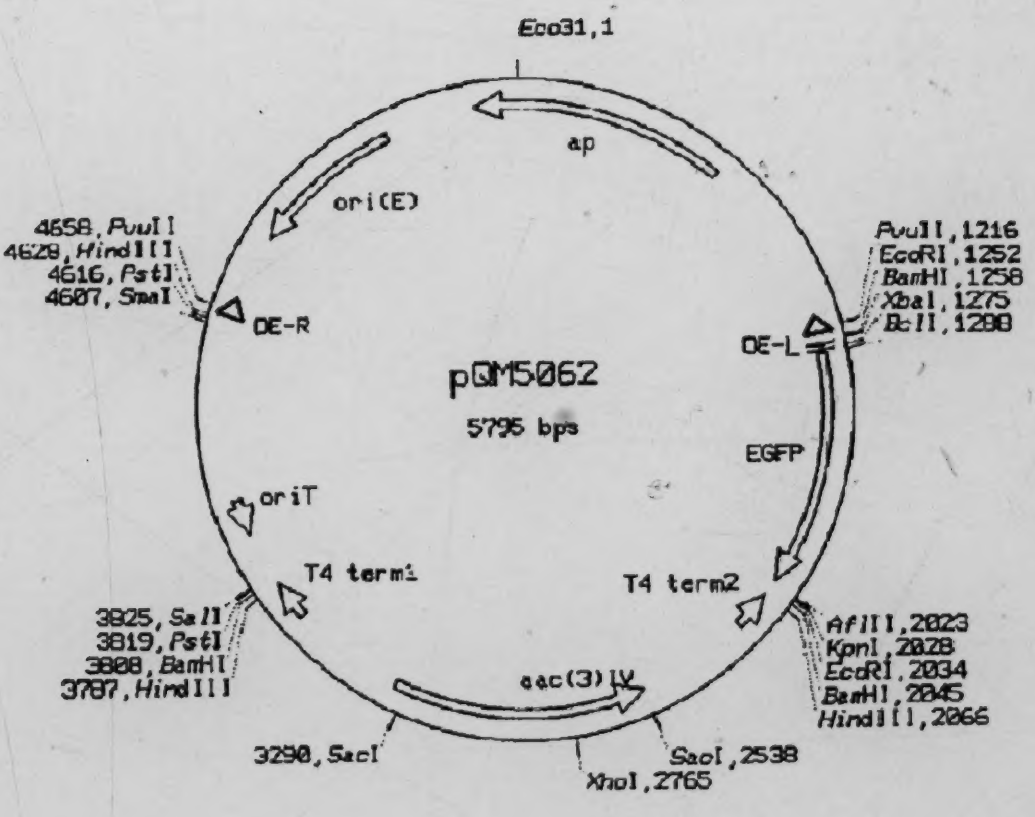
When *E. coli* cultures carried plasmids, cosmids, or transposon-containing clones with selectable antibiotic resistance markers, antibiotics were added to a final concentration of 100 μg/mL for ampicillin (Amp), 30 μg/mL for kanamycin (Km), 25 μg/mL for chloramphenicol (Cm), or 50 μg/mL for Apra.

II.2.2 Plasmids and cosmids

The Tn*5* transposon-containing plasmid pQM5062 (Figure 11) was provided by Dr. Paul Dyson from the School of Biological Sciences, University of Wales Swansea, Swansea, United Kingdom. The plasmid was transformed into *E. coli* DH5α by electroporation upon receiving, and was maintained as glycerol stock at -70°C.

The transposon-containing plasmid pQM5062 is 5795 bp in size. It has an origin for replication in *E. coli* [*ori*(E)] and harbors a 3.4 kb modified Tn*5* transposon ranging from 1216-4658 bp on the map. The transposon is flanked by two IRs (OE-L and OE-R). It carries an Apra resistance gene [*aac(3)IV*], allowing for antibiotic selection and maintenance of transposon-containing clones; an origin of transfer (*ori*T), which makes it possible for conjugal transfer of DNA between different species; and an *egfp* (enhanced green fluorescent protein) gene which will facilitate the *in vivo* transcriptional analysis of the transposon-disrupted genes and could be used to reveal the cellular location of the disrupted protein in defined manners. The Tn*5* transposon can be released from the plasmid by *Pvu*II digestion. Since the Tn*5* used in this study is different from the wildtype Tn*5*, not only in structure, but also in its lack of a transposase gene, thus it should only be considered as a "modified transposon" although the term "transposon" is used for the rest

Figure 11. Map of modified Tn*5* transposon-containing plasmid pQM5062 (provided by Dr. Paul Dyson from the School of Biological Sciences, University of Wales Swansea, Swansea, United Kingdom).


Eco31,1
ap
ori(E)
4658, PvuII
4628, HindIII
4616, PstI
4607, SmaI
OE-R
PvuII,1216
EcoRI,1252
BamHI,1258
XbaI,1275
BclI,1298
OE-L
pQM5062
5795 bps
EGFP
oriT
T4 term1
T4 term2
3825, SalI
3819, PstI
3808, BamHI
3787, HindIII
AflII,2023
KpnI,2028
EcoRI,2034
BamHI,2045
HindIII,2066
aac(3)IV
3290, SacI
XhoI,2765
SacI,2538

of this study.

The recombinant cosmid 6G9 analyzed in this study is from a cosmid library that was constructed by Mr. Wook Jin, University of Alberta, and that contained fragments of genomic DNA from *S. clavuligerus* in the cosmid vector pWE15 (Stratagene). The cosmid vector is 8.2 kb in size and contains an Amp resistance gene (*ampr*), a Km resistance gene (*neor*), the origin of replication in *E. coli* (Col E1 *ori*), the origin of replication in SV40 (SV40 *ori*) and cos sites. T3 and T7 bacteriophage promoters flank the insert. Cosmid 6G9 was originally isolated from the library based on its content of clavam biosynthetic genes.

II.3 Isolation of DNA

II.3.1 Isolation of plasmid and cosmid DNA from *E. coli*

E. coli plasmid and cosmid DNA was isolated by the alkaline lysis method (Sambrook et al. 1989). All plasmid and cosmid DNA preparations were analyzed by agarose gel electrophoresis (see section 11.5.1).

II.3.2 Preparation of genomic DNA from *Streptomyces*

Genomic DNA was isolated from *S. clavuligerus* cultures using the modified method of Kieser et al. (2000). Fifteen-milliliter amounts of mycelial cultures were collected by centrifugation, washed twice with 10.3% (w/v) sucrose and re-suspended in 5 mL of lysozyme buffer (25 mM Tris-Cl, 25 mM EDTA, 0.3M sucrose, pH 8.0, add 2-4 mg/mL lysozyme before use). After incubation at 37°C for 45 minutes with periodic gentle mixing, 2.5 mL amounts of 2% (w/v) SDS were added and mixed by inverting. Samples

were then extracted with 4 mL amounts of neutral phenol/chloroform, and centrifuged for 10 min. The aqueous layers were transferred with end-clipped pipet tips to new tubes. Neutral phenol/chloroform extraction was repeated for at least three times until no protein interface between the two phases could be seen. The aqueous layers were then extracted with 4 mL amounts of chloroform twice by the same methods. After extraction, one-tenth volume of 3M sodium acetate was added to the aqueous layers and two volumes of 95% ethanol was layered on the top. After gentle invertion and mixing, genomic DNA was collected by either centrifugation or being spooled onto a sealed Pasteur pipette and transferred into a new tube. The DNA was washed with 70% ethanol, air-dried, and dissolved in water before storage at -70°C.

II.3.3 Determining DNA Concentration and Evaluating DNA Quality

The concentration and purity of DNA preparations was measured using a UNICAM UV/Vis Spectrometer UV3 (ATI Unicam, Cambridge, UK). The spectrophotometer was blanked with the solution in which the DNA was resuspended. After adding the DNA samples into cuvettes, the samples were scanned between 220 and 320 nm. The accuracy range of the absorbance values should be between 0.1 and 0.9.

DNA concentration was calculated by using the following formula: DNA Concentration in μg/mL = (A260 Reading) x (Dilution Factor) x (50 μg/mL Constant). DNA purity was determined by the ratio of absorbance at 260 nm : absorbance at 280 nm. The ideal ratio should be 1.8-2. Higher ratios suggest contamination of RNA in the sample, while lower ratios suggest contamination of protein and phenol in the sample.

II.4 Introduction of DNA into bacterial cells

II.4.1 Transformation of DNA into *E. coli*

To prepare *E. coli* competent cells for electroporation, an overnight *E. coli* culture was used to inoculate fresh LB broth at a 1% (v/v) inoculum. The culture was then incubated on a 37°C shaker at 250 rpm until an OD_{600} of 0.35-0.5 was reached. The cells were harvested by centrifugation at 4°C for 10 min at 4000-8000 rpm in a Sorvall RC2-B automatic superspeed refrigerated centrifuge (Norwalk, CT) and the supernatant was decanted. After washing twice with cold sterile 10% (w/v) glycerol, first with an equal volume and then with a half volume, the cells were resuspended in 1-2 mL of cold sterile 10% (w/v) glycerol, and cell suspensions were dispensed into chilled 1.5 mL microfuge tubes at 40 µl per tube. Cells were then rapidly frozen in a dry-ice/ethanol bath and stored at –70°C until use.

Before electroporation, DNA and 2 mm gap electroporation cuvettes (BTX, San diego, CA) were thoroughly chilled on ice. The electroporation chamber (Bio-Rad Laboratories Ltd., Hercules, CA) was also thoroughly chilled to 4°C, and *E. coli* competent cells were slowly thawed on ice. DNA was added to thawed electrocompetent *E. coli* cells, and was mixed with the cells by flicking the tube gently. The cell/DNA mixture was incubated on ice for 1 min before being transferred into the chilled cuvette. The cuvette was then placed into the electroporation chamber, and pulsed at a setting of 25 microfarads, 200 ohms resistance and 2.5 kilovolts in a Gene Pulser II electroporation apparatus (Bio-Rad). Immediately after pulsing, 950 µl amounts of LB broth were added.

The cell suspension in the cuvette was mixed by gently pipetting up and down, transferred to a new tube, and incubated at 37°C for 1 hr, allowing the cells to recover. After recovery, electroporated cells were plated as serial dilutions onto LB agar plates supplemented with antibiotics appropriate for selection of transformants. The plates were incubated at 37°C for 14-18 hr.

II.4.2 Conjugation of DNA into *S. clavuligerus*

Transposon-containing cosmid DNA was introduced into *S. clavuligerus* using a modification of the method of Kieser et al. (2000). *E. coli* ET12567 (pUZ8002) was used as a non-methylating donor for conjugation. The *E. coli* ET12567 (pUZ8002) cells were grown in the presence of Km and Cm to maintain selection for pUZ8002 and the *dam* mutation. Electrocompetent cells of *E. coli* ET12567 (pUZ8002) were prepared as described above. The competent cells of *E. coli* ET12567 (pUZ8002) were transformed with the transposon-containing cosmids by electroporation, and the transformed cells were selected on LB agar plates supplemented with Cm, Km and Apra.

To conjugate transposon-containing cosmids from *E. coli* to *S. clavuligerus* 3585, one isolated transformant was used to inoculate a 5 mL amount of LB broth supplemented with Cm, Km and Apra. The culture was incubated at 37°C on a tube roller for 14-18 hr. The culture was then used to inoculate 5 mL of LB broth supplemented with Cm, Km and Apra at a 1% (v/v) innoculum, and incubated at 37°C until reaching an OD_{600} of 0.4-0.6. The cells were harvested by centrifugation and washed twice with equal volumes of LB broth, and resuspended in 0.1 volume (500 µl) of LB broth. To a 500 µl amount of washed

E. coli cell suspension, approximately 10^8 *S. clavuligerus* 3585 spores that had been heat-shocked at 50°C for 10 min and allowed to cool to room temperature were added. The mixture was centrifuged briefly and most of the supernatant was decanted. The pellet was resuspended in the residual liquid (200-300 μl), plated out on AS-1 (Mg) agar, and incubated at 28°C. AS-1 (Mg) agar contains: yeast extract 1 g, L-alanine 0.2 g, L-arginine 0.2 g, L-asparagine 0.5 g, soluble starch 5 g, NaCl 2.5 g, Na_2SO_4 10 g, agar 20 g, 10 mL of 1M $MgCl_2$ stock per liter, adjusted pH to 7.5. After 20-24 hr, the agar surface was overlaid with 1 mL sterile water containing 0.5 mg nalidixic acid and 0.5 mg Apra. The overlay was distributed over the surface of the plate by tilting. The plates were then incubated at 28°C for a further 5-6 days for colony development. Potential exconjugants were patched onto ISP4 agar containing Apra for further examination.

II.5 Techniques used for DNA analysis

II.5.1 Agarose gel electrophoresis

For routine DNA gel electrophoresis, 0.8% agarose gels in 1 x TAE buffer (40 mM Tris-Acetate, 1 mM EDTA, pH 8.0) were used. Electrophoresis was usually carried out at 80-100V at room temperature for 1-1.5 hr. After electrophoresis, the gels were stained in 0.5 μg/mL ethidium bromide for 10-15 min, and were visualized under UV light. The molecular weight markers used were *Bst*EII restriction fragments of λ phage DNA.

II.5.2 Purification of DNA fragments from agarose gels

QIAquick gel extraction kit was used to elute DNA from agarose gels. Restriction endonuclease-digested DNA fragments were separated by agrose gel electrophoresis. The

gels were then visualized under UV light and the DNA fragments of interest were excised from the gels. Purification of DNA fragments from the gel slice was done according to the manufacturers' instructions.

II.5.3 Southern hybridization

The procedure followed for transferring separated DNA fragments from an agarose gel to a nylon membrane was based on the protocol described by Sambrook et al. (1989). The DNA samples were separated on a 0.8% agarose gel by electrophoresis for 14-18 hr at 20-25 volts, and visualized with UV light. The gel was processed for blotting by the following steps. Between each step the gel was rinsed in distilled water: initially the gel was soaked with gentle shaking in ten volumes of 0.25M HCl for 30 min, and then DNA fragments were denatured by soaking in ten volumes of denaturing solution (1.5M NaCl, 0.5M NaOH) twice for a total of 40 min with gentle agitation. The gel was then neutralized twice by soaking in ten volumes of Neutralizing solution (1.5M NaCl, 0.5M Tris-HCl, pH 7.4) each time for a total of 40 min. Then the gel was placed onto the transfer apparatus, and DNA was passed to a Hybond-N nylon membrane (Amersham) via capillary action using 20 x SSC (3M NaCl, 0.3M $Na_3C_6H_5O_7 \cdot 2H_2O$, pH 7.0) as the transfer solution. After 14-18 hr of transfer, the nylon membrane was rinsed briefly in 2 x SSC and allowed to air-dry before baking in an 80°C vacuum oven for 2 hr to fix the nucleic acid to the membrane.

To label DNA fragments by nick-translation, one microgram of DNA restriction fragments was labeled with 10 μCi of [α-^{32}P]dCTP in a reaction mixture containing: 3 μl

10 x nick translation buffer, 1 mM dTTP, 1 mM dATP, 1 mM dGTP, 2.5 units of *E. coli* DNA polymerase I, 0.0025 unit of DNAse I, and sterile distilled water to give 30 µl. The 10 x nick translation buffer contains: 0.1M Tris pH 7.4, 0.1M $MgCl_2$, 0.1M β-mercaptoethanol, 0.5M NaCl, and 500 µg/mL bovine serum albumin (BSA). The nick-translation reaction mixture was incubated at 15°C for two hr. After incubation, a 15 µl amount of 0.5M EDTA was added and the solution was heated at 100°C for 3-4 min to stop the reaction.

Blots were hybridized using the procedure outlined in the Amersham Manual (1997) with minor alterations. Nylon membranes were placed in a screw cap hybridization tube with pre-hybridization solution containing Church buffer [0.5M Na_2HPO_4, 7% (w/v) SDS, 1mM EDTA, pH7.2] and 100 µg/mL of denatured salmon sperm DNA. Ten and twenty milliliter amounts of pre-hybridization solution were used for the small and large roller tubes, respectively. The blots were incubated at 65°C for 2 hr while rotating in a hybridization oven. Hybridization of the desired probe was performed by adding the nick-translation reaction mixture into the pre-hybridization solution. The tube was once again placed in the hybridization oven to continue incubation with rotation for 14-18 hr at 65°C.

After hybridization, the blots were removed from the tube and washed twice with 2 x SSC/0.1% SDS (w/v) at room temperature for 10 min, once with 1 x SSC/0.1% SDS at 65°C for 15 min, and once with 0.1 x SSC/0.1% SDS at 65°C for 10 min. Blots were then wrapped in plastic wrap and exposed to a PhosphorImager (Amersham) for 6-8 hr.

II.5.4 DNA sequencing and sequence analysis

A modified reaction mixture based on the DYEnamic ET Terminator Cycle Sequencing System (Amersham) was used. Cosmid 6G9 DNA was first digested with *Hin*dIII and then used as the template for sequencing. After stopping the restriction digestion by heat treatment, the DNA was precipitated by adding two volumes of 95% ethanol and 0.1 volumes of 3M sodium acetate, washed with 70% ethanol, and resuspended in sterile distilled water. Forward and reverse sequencing primers that bind near the ends of the transposon were used for bidirectional sequencing of an insertion site. They are HCA2 (Forward primer 5' CGACGAGCAAGGCAAGACCGATC 3') and HCA1 (Reverse primer 5' CGGTGAACAGCTCCTCGCCCTTG 3').

The sequencing reaction used was: 1000 ng *Hin*dIII digested 6G9 DNA, 40 pmol primer, 8 µl ET terminator sequencing reagent premix, 5% DMSO, and distilled water was added up to a total volume of 20 µl. Sequencing reaction mixtures were prepared in 250 µL thin walled PCR tubes, and were overlaid with approximately 80 µL of mineral oil. PCR tubes were then placed into a MiniCycler™ thermal cycler (MJ Research Inc., Watertown, MA). The thermal cycling program used was: denature at 96°C for 30 sec, anneal at 60°C for 120 sec, repeat for 35 cycles. The reaction mixtures were then transferred into a 1.5 mL microcentrifuge tube containing 2 µL of the supplied sodium acetate/EDTA buffer and precipitated by the addition of 80 µL of 95% ethanol. After incubation on ice for 15 min, the reaction product was centrifuged at 14000 rpm for 15 min in a centrifuge (model 5415C, Brinkmann Instruments. Inc. Germany) to sediment

DNA. After washing with 200 µL of 70% ethanol and air-drying, the pellet was sent to the DNA Synthesis Laboratory at the Department of Biological Sciences, University of Alberta, for sequencing analysis.

The original DNA sequencing chromatograms obtained were analyzed using GeneTools version 1.0. Sequence overlaps were detected by the Assembly Editor of GeneTools. Putative ORFs were examined by Frameplot 3.0 beta (Ishikawa and Hotta 1999) by calculating biased usage of G and C nucleotides in the third position of codons and detecting start and stop codons nearby the ORFs. Similarity searches were accomplished using the online BLAST program at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/BLAST/). To locate functional domains, an NCBI CDD search, which compared protein sequences with the conserved domain databases, was used. A Kyte-Doolittle hydropathy plot (Kyte and Doolittle 1982) was used to predict the location of putative proteins within the cell. The number of amino acids and the calculated molecular weights of the putative proteins were determined using the DNA Strider 1.2 program (Douglas 1995).

II.6 Generation of DNA probe fragments by PCR

DNA probe fragments were generated by PCR amplification. The probes were used to determine the appropriate *Eco*RI fragment of cosmid 6G9, which was the candidate for this study.

Primers used for amplification of the probes were HCA6 (Forward primer 5' CCGGGGGCGAGGACGACT 3') and HCA5 (Reverse primer 5'

CCGTTGTCCAGCGCTCGTTCGT 3'). The PCR reaction mixture used was: 0.5 μM amounts of each primer, 1000 ng *Eco*RI and *Hin*dIII digested cosmid 6G9 as template DNA, 10 μl 10 x PCR buffer (Boehringer), 5 μl 4 mmole/L dNTP, 3 μl DMSO, and distilled water was added to a final volume of 100 μl. The PCR reaction mixtures were prepared in 250 μL thin walled PCR tubes, and were overlaid with approximately 80 μL of mineral oil. PCR tubes were then placed into a MiniCycler™ thermal cycler. After an initial 5-min denaturation step at 94°C, 2 units of *Taq* DNA Polymerase were added to the reaction mixture and the PCR reaction was continued with the following conditions: 94°C for 30 sec, 60°C for 45 sec, and 68°C for 2 min, repeat for 30 cycles, followed by a final extension step at 72°C for 7 min.

The PCR products were examined by electrophoresis, and were eluted from agarose by using a QIAquick gel extraction kit. The fragment was then radioactively labeled with ^{32}P and used as the probe for Southern hybridization.

II.7 Construction of *S. clavuligerus* mutants

II.7.1 *In vitro* transposon mutagenesis

The modified Tn*5* was released from the transposon-containing plasmid pQM5062 (Figure 11) by *Pvu*II digestion, and was separated from the rest of the plasmid by electrophoresis. After examining the ethidium bromide stained gel under ultraviolet light, the gel slice containing the 3.4 kb Tn*5* fragment was cut from the agarose, and the DNA was purified using a Qiaquick gel extraction kit. The target DNA for the *in vitro* transposition reaction, cosmid 6G9, was prepared to be free of contaminating

chromosomal DNA which would act as a direct competitor for transposon insertion.

The *in vitro* transposition reaction mixture was set up, incubated and stopped according to the supplier's specifications (Epicentre, Madison WI, Figure 12). For a 10 µl transposition reaction system, 0.2 µg of cosmid 6G9 was combined with an equivalent amount of transposon DNA on a molar basis in the presence of 1 µl of EZ::TN transposase, 1 µl of 10 x reaction buffer, and sterile water to make up a final volume of 10 µl. The following formula was used to calculate the molar amount of DNA used in the transposition reaction: µmole DNA = µg DNA/[(number of base pairs in target DNA) x 600]. The reaction mixture was incubated at 37°C for 2 hr and was stopped by adding 1 µl EZ::TN 10 x Stop Solution and heated at 70°C for 10 min.

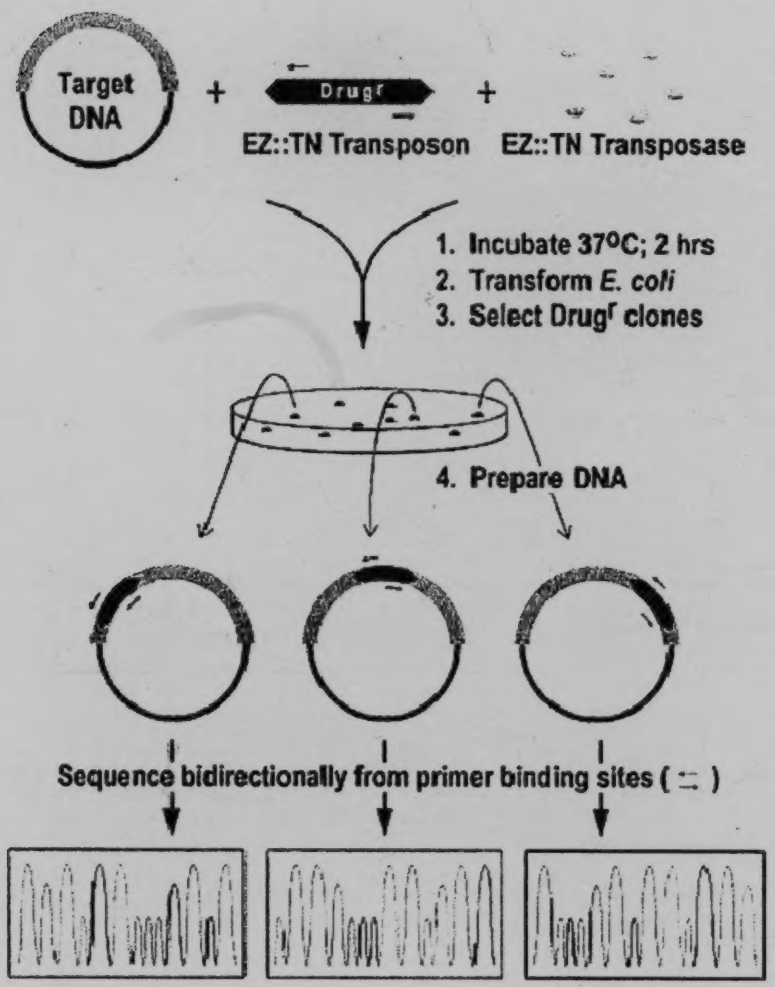
One micro liter of reaction mixture was used for electroporation into TransforMax™ EC100™ Electrocompetent *E. coli* cells and the transformed cells were plated on LB agar supplemented with Apra. The plates were incubated at 37°C for 14-18 hr. The transformants were maintained as glycerol stocks in 96-well microplates at -70°C.

II.7.2 Construction of *S. clavuligerus* mutant strains

Based on DNA sequence analysis, several putative ORFs were identified. At least one transposon-containing cosmid from each putative ORF was chosen to construct *S. clavuligerus* mutant strains. Priorities were given to those that lay in the same orientation with the *egfp* gene and were located from the 5' end to the middle region of the ORFs.

To introduce transposon-containing cosmids into *S. clavuligerus* 3585, DNA from mutant clones was prepared and then electroporated into *E. coli* ET12567 (pUZ8002).

Figure 12. *In vitro* transposon mutagenesis reaction (adopted from the protocol of Epicentre). The reaction mixture containing cosmid DNA 6G9, released modified Tn*5* transposon, and transposase was incubated at 37°C for 2 hr and the reaction was then stopped. After electroporation, transformants were selected with apramycin. DNA of transposon-containing clones was prepared and used as template for sequencing the transposon-flanking regions.


Target DNA
+
Drug^r
+
EZ::TN Transposon
EZ::TN Transposase
1. Incubate 37°C; 2 hrs
2. Transform *E. coli*
3. Select Drug^r clones
4. Prepare DNA
Sequence bidirectionally from primer binding sites (=)

These *E.coli* strains then served as the donor cells for conjugation of the tranpsoson-insertion cosmids into *S. clavuligerus* 3585 as described above (see section II.4.2). After development of potential exconjugants on AS-1 (Mg) plates containing Apra and nalidixic acid, three exconjugants of each mutant clone were patched onto ISP4 plates supplemented with Apra. Spores of each exconjugant were harvested from patches, collected in microcentrifuge tubes containing sterile water, dispersed by Branson B-42 ultrasonic cleaner (Shelton, CT) for five min, and filtered through sterile non-absorbent cotton packed into a 10 mL syringe. After filtration, spore suspensions were serially diluted in sterile water, and plated on ISP4 agar plates supplemented with Apra to allow for the development of isolated single colonies. Once sporulated, spores from single colonies were used to inoculate ISP4 plates to generate sufficient spores for further study and preparation of glycerol stocks of spores.

To screen for double-crossovers resulting in gene replacement, spores of *S. clavuligerus* mutant strains were used to inoculate ISP4 agar plates supplemented with Km (conferred by pWE15). No growth after ten days indicated occurrence of double-crossovers between the transposon-containing cosmids and the chromosomal DNA of *S. clavuligerus* 3585. Genomic DNA from the *S. clavuligerus* mutant strains constructed was prepared and analyzed by Southern hybridization to confirm the existence of the transposon on the chromosome.

II.8 β-lactam bioassay and HPLC analysis

Both *S. clavuligerus* 3585 and the mutant strains constructed in this study were

cultivated in TSBS as seed cultures for 40 hr before subculture into Soy medium as described above (see section II.2.1). Soy cultures were sampled at 72 and 96 hr. Cultures were harvested by centrifugation and the supernatants were used for β-lactam bioassay and HPLC analysis.

II.8.1 β-lactam bioassay

Bioassays were conducted by the agar plate diffusion method. Molten agar was inoculated with indicator organism and then poured into plates. Twenty-five microliters of culture supernatant of each strain was applied to a filter disk set on the agar. After incubation for about 20 hr, zones of inhibition were measured in millimeters as a function of the amount of β-lactam antibiotics in the culture supernatant.

For cephamycin C bioassay, the amount of total antibiotics produced was assayed by using *E. coli* ESS as the indicator organism seeded into TSB agar. After 24-hr growth at 37°C, zones of inhibition were measured as a function of the amount of β-lactam antibiotics in the culture supernatant.

For clavulanic acid bioassay, β-lactamase inhibitory activity was assayed by using *K. pneumoniae* as the indicator organism seeded into TSB agar containing a final concentration of 6 μg/mL filter-sterilized penicillin G. Control plates lacked penicillin G. After 24-hr growth at 37°C, zones of inhibition were measured as a function of the amount of clavulanic acid in the culture supernatant.

For alanylclavam bioassay, the anti-metabolite activity of alanyl-clavam was determined with *Bacillus* spp. ATCC 27860 as the indicator organism. Samples of culture

supernatant were applied to filter disks on the surface of Davis minimal media inoculated with *Bacillus* sp. ATCC 27860. Davis minimal media contains: K_2HPO_4 7 g, KH_2PO_4 3 g, Na_3citrate·$2H_2O$ 0.5 g, $MgSO_4 \cdot 7H_2O$ 0.1 g, $(NH_4)_2SO_4$ 1 g, agar 18 g per liter adjusted to pH7.0. After autoclaving, 20% (w/v) of glucose was added to give a final glucose concentration of 0.2% (v/v). Control plates contained filter-sterilized L-methionine at 200 μg/mL to counter act the antimetabolite activity of alanyl-clavam. Zones of inhibition were measured after 14-18 hr growth at 28°C.

II.8.2 Analysis of clavulanic acid and other clavam metabolites by HPLC

II.8.2.1 Analysis of clavams in liquid cultures

The levels of clavulanic acid and *5S* clavam metabolites were analyzed by HPLC as described before (Jensen et al. 2000). The HPLC analysis was performed with an Alliance HPLC equipped with a photo diode array detector, all from Millipore Waters (Mississauga, ON). Culture supernatants were centrifuged to remove particulate matter, and then a 100 μl sample was mixed with 25 μl imidazole reagent that reacts specifically with β-lactam-containing compounds, and incubated at room temperature for 15 min to form imidazole-derivatized clavams. A 50 μl sample of the derivatized supernatant was then analyzed on a reverse-phase column (Bondapak-C18) with an isocratic buffer system consisting of 0.1 M KH_2PO_4-6% methanol (pH 3.68; adjusted with glacial acetic acid). Peaks representing clavulanic acid and clavam-2-carboxylate in culture broths of *S. clavuligerus* were identified by comparison with authentic standards. 2-Hydroxymethylclavam was identified by its known chromatographic properties relative

to those of clavulanic acid and clavam-2-carboxylate.

Later analyses were conducted on the same instrument but equipped with a reverse phase column (Xterra 0.21 x 10 cm, Waters Scientific, Milford, MA). Imidazole-derivatized culture supernatants (5 µl) were analyzed at a flow rate of 0.25 ml/min. The mobile phase consisted of solvent A (10 mM ammonium bicarbonate, pH 10) and solvent B (acetonitrile), used in a gradient system as follows: 100% solvent A for 5 min, linear gradient to 85% solvent A over 10 min, 85% solvent A for 5 min, linear gradient to 100% solvent A over 1 min, 100% solvent A for 9 min. Eluant was monitored at 311 nm to detect the imidazole derivatives.

II.8.2.2 Analysis of clavams in solid cultures

Glycerol spore stocks of both *S. clavuligerus* 3585 and the mutant strains constructed in this study were inoculated on soy agar plates to give patches with a diameter of about 1.5 centimeters. To avoid antibiotic diffusion from one patch to the other, fifteen-centimeter petri dishes were used and the soy agar in each petri dish was cut into three equal parts, inoculated with three exconjugants of each mutant separately.

After incubation at 28°C for six days, each mutant patch was transferred into a 3ml syringe, and frozen at -70°C for 14-18 hr. The liquid in each syringe was squeezed out of the partially thawed agar and collected into a new microfuge tube the next day. After a brief centrifugation to remove any mycelia or particulate matter, the aqueous samples were tranferred into new tubes and were analyzed by HPLC.

III. RESULTS

The details of the biosynthetic pathway leading from clavaminic acid to the various clavam metabolites in *S. clavuligerus* are essentially unknown. The fact that *cvm4* and *cvm5* disruption mutants fail to produce clavam-2-carboxylate, 2-hydroxymethylclavam and alanylclavam suggests that a common precursor may serve as a branched-point intermediate to all the clavam metabolites (Mosher et al. 1999), but the enzymatic functions of these newly discovered genes are unclear. Moreover, the enzymes that catalyze the biosynthesis of any one specific *5S* clavam metabolite have not been identified yet. Therefore, identification and characterization of more putative *5S* clavam biosynthetic genes are one of the main goals of the research within the Jensen group. Since the genes encoding enzymes for antibiotic biosynthesis are clustered on the chromosomes of bacterial antibiotic producers and *S. clavuligerus* is no exception, the open reading frames surrounding the newly uncovered paralogue gene cluster seem to be good candidates for study.

In this study, the genetic region of the *ceaS1* side of the paralogue gene cluster was analyzed. In particular, the Tn*5 in vitro* transposon mutagenesis system was applied on cosmid 6G9 to create gene mutants. By introduction of transposon-containing cosmids into wildtype *S. clavuligerus*, mutant strains were constructed. The involvement of the disrupted genes in clavam biosynthesis in *S. clavuligerus* was then assessed by bioassay and HPLC.

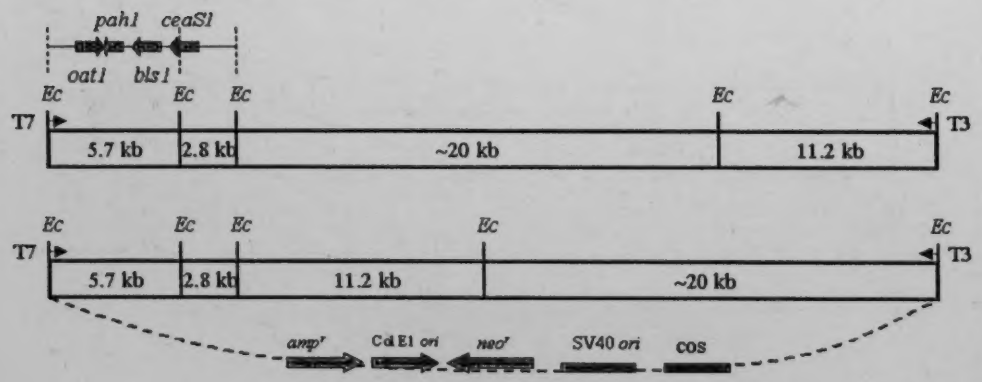
III.1 Identifying the *Eco*RI fragment of interest on cosmid 6G9

Previous restriction analysis by the Jensen group showed that the *S. clavuligerus* DNA insert on cosmid 6G9 was approximately 40 kb in size (Figure 13). Three *Eco*RI restriction sites were found on the insert, such that digestion with *Eco*RI was able to separate the insert into a 5.7 kb, 2.8 kb, 11 kb and a fragment of approximately 20 (~20) kb in size. Two additional *Eco*RI restriction sites were present in the multiple cloning site of the cosmid vector, pWE15 (8.2 kb), which could release the pWE15 from the *S. clavuligerus* DNA insert.

Several open reading frames had already been identified on cosmid 6G9. This includes four genes, *ceaS1*, *bls1*, *pah1*, and *oat1*, from the paralogue gene cluster and an uncharacterized gene, *orfA*. Southern hybridization and restriction analysis showed that the 5' end of the 2.8 kb *Eco*RI fragment of cosmid 6G9 contained the 5' end of *ceaS1*; while the 3' end of *ceaS1*, together with the *bls1*, *pah1*, and *oat1* genes, was present in the 5.7 kb *Eco*RI fragment of cosmid 6G9 (Figure 13; Jensen et al. 2003b; Tahlan et al. 2003). The *orfA* gene was present in the 3' end of the 2.8 kb *Eco*RI fragment (Jensen et al. unpublished data), but sequencing of *orfA* was incomplete and its putative involvement in the biosynthesis of clavams in *S. clavuligerus* has not been characterized yet. Furthermore, the 5' end of the 5.7 kb *Eco*RI fragment was known to lie at the T7 end of the *S. clavuligerus* DNA insert.

The candidate for this study was the *Eco*RI fragment that lay next to the 3' end of the 2.8 kb *Eco*RI fragment. However, the order of the 11 kb and the ~20 kb *Eco*RI fragments on cosmid 6G9 was not clear. Therefore, the first step of this study was to determine the

Figure 13. Partial map of cosmid 6G9. The *Eco*RI restriction sites (*Ec*) are indicated. Two possible arrangements of the 11 kb and 20 kb *Eco*RI fragments on 6G9 are shown. The paralogue genes are labeled. The cosmid vector pWE15 is indicated as dashed line and the drawing of the vector is not to scale. *amp*^{r}: ampicillin-resistance gene; Col E1 *ori*: the origin of replication in *E. coli*; *neo*^{r}: kanamycin-resistance gene; SV40 *ori*: the origin of replication in SV40; cos: the cos sites. T3 and T7 bacteriophage promoters flank the insert.


pahl
ceaSl
oatl
blsl
Ec
Ec
Ec
Ec
Ec
T7
T3
5.7 kb
2.8 kb
~20 kb
11.2 kb
Ec
Ec
Ec
Ec
Ec
T7
T3
5.7 kb
2.8 kb
11.2 kb
~20 kb
amp'
ColE1 ori
neo'
SV40 ori
cos

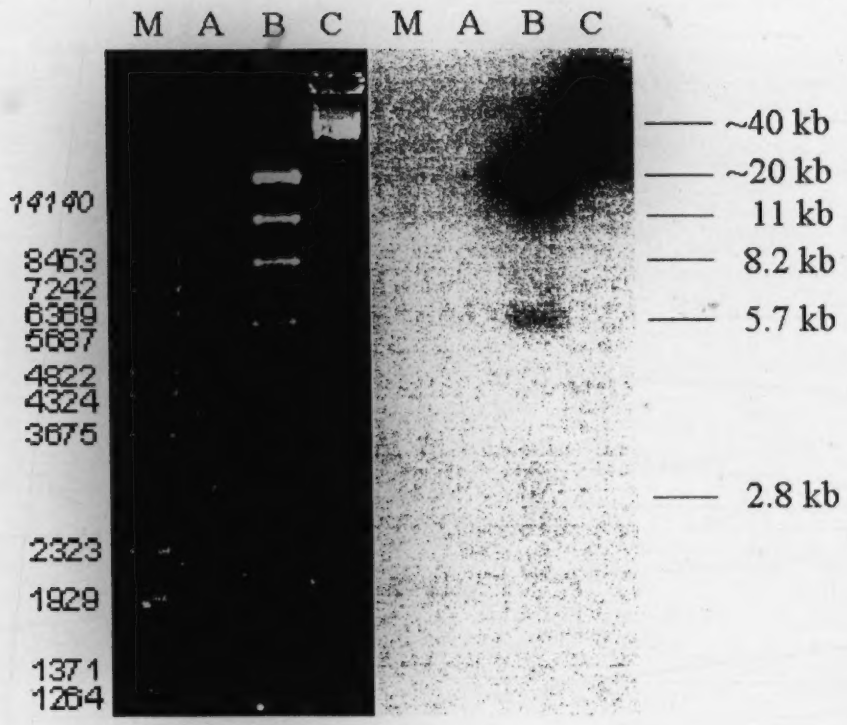
pWE15 8.2 kb

order of the 11 kb and ~20 kb *Eco*RI fragments on cosmid 6G9. To locate these two *Eco*RI fragments, PCR amplification and Southern hybridization were conducted. Since the sequencing of the 3' end of the 2.8 kb *Eco*RI fragment was incomplete, the sequence of the T3 end of pWE15 was used to design PCR primers to locate the *Eco*RI fragment that lay next to the T3 end of pWE15.

A 21-base primer, HCA3 (5' CGGCCGCAATTAACCCTCACT 3'), was designed based on the sequence of the T3 end of pWE15, and was used to sequence across the junction of pWE15 and into the insert of cosmid 6G9. An approximate 600 bp sequence of cosmid insert was obtained (data not shown), and was then used to design a pair of PCR primers, which would be able to amplify a 517 bp DNA fragment of cosmid insert. They were HCA6 (Forward primer 5' CCGGGGGCGAGGACGACT 3') and HCA5 (Reverse primer 5' CCGTTGTCCAGCGCTCGTTCGT 3'). After PCR amplification by using HCA5 and HCA6 as primers, the PCR product was examined by electrophoresis, eluted from agarose, radioactively labeled with ^{32}P, and finally used as a probe for Southern hybridization analysis.

To locate the 11 kb and ~20 kb *Eco*RI fragments on cosmid 6G9, DNA of both *Eco*RI-digested cosmid 6G9 and uncut cosmid 6G9 was separated by agarose gel electrophoresis, followed by transfer to a nylon membrane. The membrane was then probed with the ^{32}P-labeled 517-bp PCR product. After exposure to PhosphorImager, it was shown that the 517 bp PCR product hybridized to the uncut cosmid 6G9 DNA, a positive control, and to the ~20 kb *Eco*RI fragment of cosmid 6G9 (Figure 14). A very

Figure 14. Southern hybridization analysis of *Eco*RI fragments of cosmid 6G9. Left, Ethidium bromide stained agarose gel used for Southern hybridization. Right, Southern hybridization blot probed with the 517 bp PCR product amplified from the *Eco*RI fragment that lies next to the T3 end of pWE15. Lane M, lambda/*Bst*EII marker. The cohesive ends of the *cos* site of bacteriophage lambda from fragments 8453 bp and 5687 bp may anneal and form an additional band at 14140 bp. Lane A, empty. Lane B, 6G9 digested with *Eco*RI. Lane C, uncut 6G9.


M A B C M A B C
14140
8453
7242
6369
5687
4822
4324
3675
2323
1929
1371
1264
~40 kb
~20 kb
11 kb
8.2 kb
5.7 kb
2.8 kb

faint hybridizing band at the 5.7 kb *Eco*RI fragment region was also seen. This band may just be a non-specific hybridization product although the possibility of additional paralogues cannot be ruled out. As a result, the ~20 kb *Eco*RI fragment of cosmid 6G9 would be the one adjacent to pWE15, and the 11 kb *Eco*RI fragment would be the one that lies next to the 2.8 kb *Eco*RI fragment, and thus the desired candidate for this study.

III.2 *In vitro* transposon mutagenesis

To analyze the *ceaS1* side of the genomic region of the paralogue gene cluster, the Tn*5*-based *in vitro* transposon mutagenesis system was applied on cosmid 6G9. This system was developed by Epicentre and modified for use in *Streptomyces* spp. by Dr. Paul Dyson, University of Wales Swansea. The Tn*5* transposon is 3.4 kb in size and carried by a 5.8 kb plasmid pQM5062. The transposon contains an origin for replication in *E. coli*, an apra-resistance gene, an origin of transfer (*ori*T), and an *egfp* gene.

The *in vitro* transposition reaction was conducted as described above (see section II.7.1). In brief, the released Tn*5* transposon was reacted with an equal amount of cosmid 6G9 on a molar basis in the presence of transposase for two hr. After stopping the reaction, one micro liter of reaction mixture was used for electroporation into TransforMax™ EC100™ Electrocompetent *E. coli* cells. The transformed cells were selected with Apra, the antibiotic resistance conferred by Tn*5*.

Thousands of Apra-resistant transformants arose from the transformation. Among them, nine hundred and sixty were randomly picked up and preserved as glycerol stocks in ten 96-well microplates at -70°C.

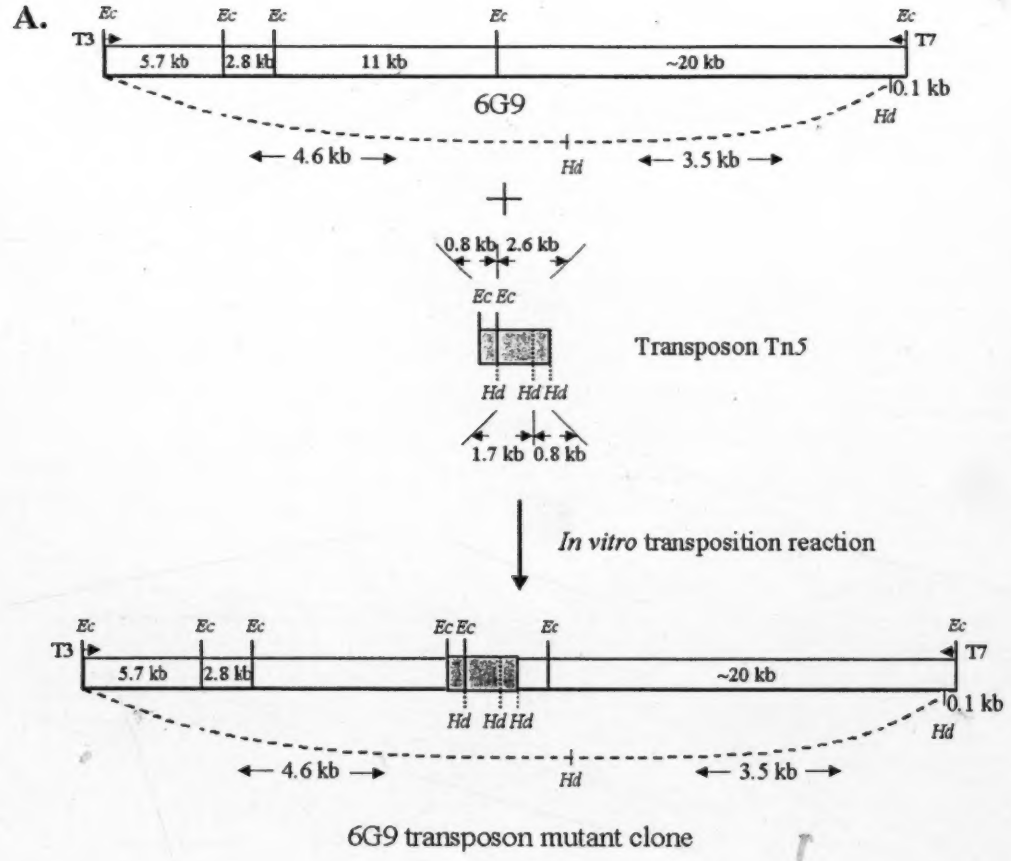
III.3 Analysis of transposon-containing clones in *E. coli*

III.3.1 Screening of transposon-containing clones

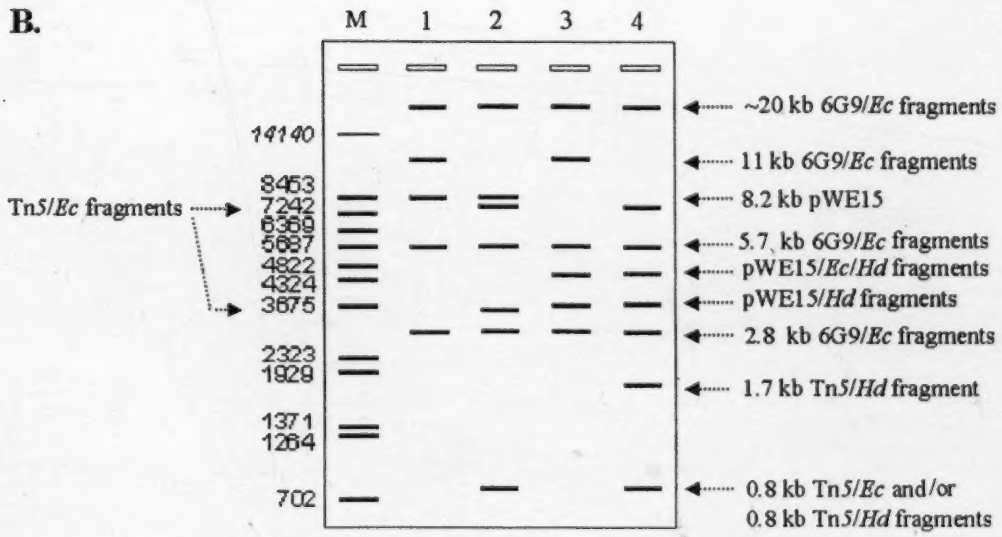
To identify desired transposon-containing clones that contained the Tn*5* inserted into the 11 kb *Eco*RI fragment, DNA of 156 randomly selected transposon-containing clones was prepared and analyzed by restriction digestion followed by agarose gel electrophoresis. After electrophoresis, the Tn*5*-containing *Eco*RI fragment was determined by observing disappearance of the original *Eco*RI fragment and the generation of new smaller fragments, due to the presence of two *Eco*RI restriction sites on the transposon. The restriction map of one of the transposon-containing clones is shown in Figure 15A as an example.

*Eco*RI single digestion was used for the first round of analysis, fragmenting the transposon-containing *Eco*RI fragment into two pieces and generating a 0.8 kb transposon DNA fragment which was used as a hallmark for the occurrence of transposition events. After electrophoresis, the disappearance of one of the *Eco*RI fragments was used to identify the transposon-containing *Eco*RI fragment. The original target *Eco*RI fragment would be cleaved into several pieces which sum up in size to that of the original fragment (Figure 15B lane 1 and lane 2). In some cases, little change in size could be observed due to the close proximity of the transposon insertion site to an *Eco*RI site. In other cases, changes in size due to insertion of the transposon into a fragment larger than 10 kb was difficult to recognize using conventional gel electrophoresis. A second round of screening was then conducted using *Eco*RI and *Hin*dIII double digestion, which completely

Figure 15. Determination of the transposon-containing *Eco*RI fragments by restriction digestion. A. Restriction map of one of the transposon-containing clones. The *Eco*RI (*Ec*) and *Hin*dIII (*Hd*) restriction sites are labeled. The transposon is shown as a grey bar. The cosmid vector is indicated by a dashed line and the drawing of the vector is not to scale. B. Diagrammatic representation of the agarose gel electrophoresis pattern of *Eco*RI-digested and *Eco*RI/*Hin*dIII-digested 6G9 and transposon-containing cosmid of panel A. Lane M, lambda/*Bst*EII marker. Lane 1, 6G9 digested with *Eco*RI. Lane 2, transposon-containing cosmid digested with *Eco*RI. Lane 3, 6G9 digested with *Hin*dIII & *Eco*RI. Lane 4, transposon-containing cosmid digested with *Hin*dIII & *Eco*RI.



6G9 transposon mutant clone



released the transposon from the fragmented target *Eco*RI fragment and therefore resulted in a more apparent change in size of the target *Eco*RI fragment (Fig. 15B lane 3 and lane 4).

The results of screening 156 transposon-containing clones are shown in Table 2. Of the 156 transposon-containing clones analyzed, 68 were found to contain Tn*5* on the ~20 kb *Eco*RI fragment of cosmid 6G9, 30 were found to contain Tn*5* on the 11 kb *Eco*RI fragment of cosmid 6G9, 16 were found to contain Tn*5* on the 8.2 kb cosmid vector, 14 were found to contain Tn*5* on the 5.7 kb *Eco*RI fragment of cosmid 6G9, and 6 were found to contain Tn*5* on the 2.8 kb *Eco*RI fragment of cosmid 6G9. No changes in size could be observed for 18 transposon-containing clones even after *Eco*RI and *Hin*dIII double digestion, and thus their Tn*5* insertion sites on different *Eco*RI fragments could not be identified.

There were two Apra-resistant clones which generated only the 5 kb and 0.8 kb bands after *Eco*RI digestion and electrophoresis. The size of the *Eco*RI fragments implied that these two clones may contain the transposon-containing plasmid pQM5062 only. These clones most likely represent contaminants which arose as a result of incomplete separation of the 3.4 kb Tn*5* from pQM5062 by restriction digestion, electrophoresis and gel extraction, and they presumably carry the intact pQM5062 transposon-containing plasmid which can confer Apra resistance to the *E. coli* cells.

Furthermore, two transposon-containing clones were found to contain double Tn*5* insertions on the same target replicon. This double insertion was verified by observing the

Table 2. Frequencies of transposon insertion into the different *Eco*RI fragments of cosmid 6G9.

Percentage (%)	*Eco*RI fragments of cosmid 6G9				
	~20 kb	11 kb	8.2 kb	5.7 kb	2.8 kb
Frequency of Tn*5* insertion [ab]	50.7 (68)	22.4 (30)	12.0 (16)	10.4 (14)	4.5 (6)
Fragment size as a proportion of cosmid 6G9	41.9	23.1	17.2	11.9	5.9

[a.] The results were calculated out of the 134 transposon-containing cosmids identified.

[b.] The number of transposon-containing cosmids that contained Tn*5* on the specific *Eco*RI fragment is shown in brackets.

disappearance of two *Eco*RI fragments after electrophoresis.

As shown in Table 2, for the 134 transposon-containing cosmids analyzed in this study, a correlation was observed between the frequencies of the transposon inserting into different *Eco*RI fragments and the proportions of the *Eco*RI fragments to cosmid 6G9 in size. These results provided preliminary evidence that the Tn*5* insertion into cosmid 6G9 was random.

III.3.2 Sequence analysis of the 11 kb *Eco*RI fragment on cosmid 6G9

After identifying the transposon-containing cosmids that contained Tn*5* on the 11 kb *Eco*RI fragment of cosmid 6G9, the transposon flanking regions were sequenced bi-directionally, thereby determining the target sites and orientation of the transposon insertions. The DRs, a 9-bp duplication generated from the target sequences as a result of the transposition, were used to join the forward and reverse sequences of each transposon-containing cosmid. The joined sequences were then lined up by examining the overlapping sequences using the Assembly Editor of GeneTools 1.0. As a result, approximately 10 kb of DNA sequence that lacked any *Eco*RI sites was obtained. To obtain sequence information to complete the 5' and 3' ends of the 11 kb *Eco*RI fragment, oligonucleotide primers (Table 3) were designed from the newly acquired sequence data and used for primer walking until reaching the *Eco*RI sites. In some regions of the 11 kb fragment, sequence information from the transposon clones was obtained for only one strand. In these cases, new primers (Table 3) were also designed from new sequence data acquired and used for sequencing of the opposite strand to minimize sequencing errors. A

Table 3. Oligonucleotide primers designed from new sequence data acquired.

Primer	5' to 3'
183F	GCGCCAAGCCGGGTGCAAGAAG
651R	GGCGCCCCGGTACAGCTCGAAGG
1002F	CGGGGTGGCGGCTCGGCTACTG
1961R	CGGCCGGGTCGTCGAAGGGATGC
2213F	GGCGGCGGACGGACTCAACG
2886R	CGCCCTGGGGACGTTCGCCTTC
4243F	CGCGCACACCATGACGACGGACT
9468F	CGGCGCGGACTCGTCGTGATAG
8198F	CCGCGGTGGCCTCCTGCTGCTCT
5226R	GGCGGTGGTGGCGTTCGGTGTC
7900F	CCGCACGTCCTGGCAGCTC
8800R	TACGCGAACGGTGTGCAGAAG
9488F	CCGCAGATGGGCCTTTGACCG
2.8R	CCGAACGGGCCGCGAGGAAC
2.8F2	GACCGCCAGGACATCGATCTCG
9.8F	TTGCGCCGCTCCTCGACAGAT
9488F2	GCTGCCAGGTCTTTCACGGTCT
B42RF	GCGGATGTTTCTCGTGCCCAATG
1R	CGCGCACCTCCCGGCACCAC
Fend	GCGGCATCGTCTTCATCGTTG
R800	GGAGAGACCATCGGCTACATC
R600	TGGCTGACGGCTGATATCTGAT
R400	TCGACAACGGCGCTCTGATCA
R200	GGACGTAGCTGTGATCGAAGA

diagram showing how the 11.4 kb sequence was acquired is shown in Figure 16.

The entire length of the DNA sequence obtained was 11443 bp with two *Eco*RI sites present at position 196 and position 11251, respectively. As a result, the exact length of the 11 kb *Eco*RI fragment on cosmid 6G9 was 11054 bp (from position 197 to position 11250). Of the 11.4 kb sequence obtained, 196 bp belonged to the 2.8 kb *Eco*RI fragment (from position 1 to position 196), and 193 bp belonged to the ~20 kb *Eco*RI fragment (from position 11251 to position 11443). The compiled 11.4 kb of sequence data obtained is shown in Figure 17a, while the amino acid sequences for each identified *orfs* are shown in Figure 17b.

The organization of the open reading frames was determined using the computer analysis by FramePlot 3.0 beta, which identifies ORFs on the basis of the characteristic third-position GC bias typical of *Streptomyces* (Wright and Bibb 1992). The frame analysis revealed the presence of one incomplete ORF and ten complete ORFs (Figure 18), designated from *orfB* to *orfL*. The incomplete ORF (*orfB*) contained an *Eco*RI site (position 196) within the ORF, and the other part of the *orfB* gene was present on the 2.8 kb *Eco*RI fragment on cosmid 6G9.

Streptomyces are known for their high overall average G + C contents (68-79%), and biased G + C frequencies in the codon's third position of ORFs (70-98%, Wright and Bibb 1992). FramePlot program predicts protein-coding regions in *Streptomyces* DNA based on this biased G and C codon preference in the third position of the ORFs. In this study, the overall average G + C content of the ORFs is 72.0% (Figure 18), and the G + C

Figure 16. Sequence data obtained shown in original contigs. Contigs are shown as small arrows. For sequences obtained from transposon-containing cosmids by using transposon-derived primers, the arrows are labeled with the name of cosmid, and an "R" is labeled for those obtained from the reverse primers. For sequences obtained from newly designed oligonucleotide primers, the arrows are labeled with the name of primers.

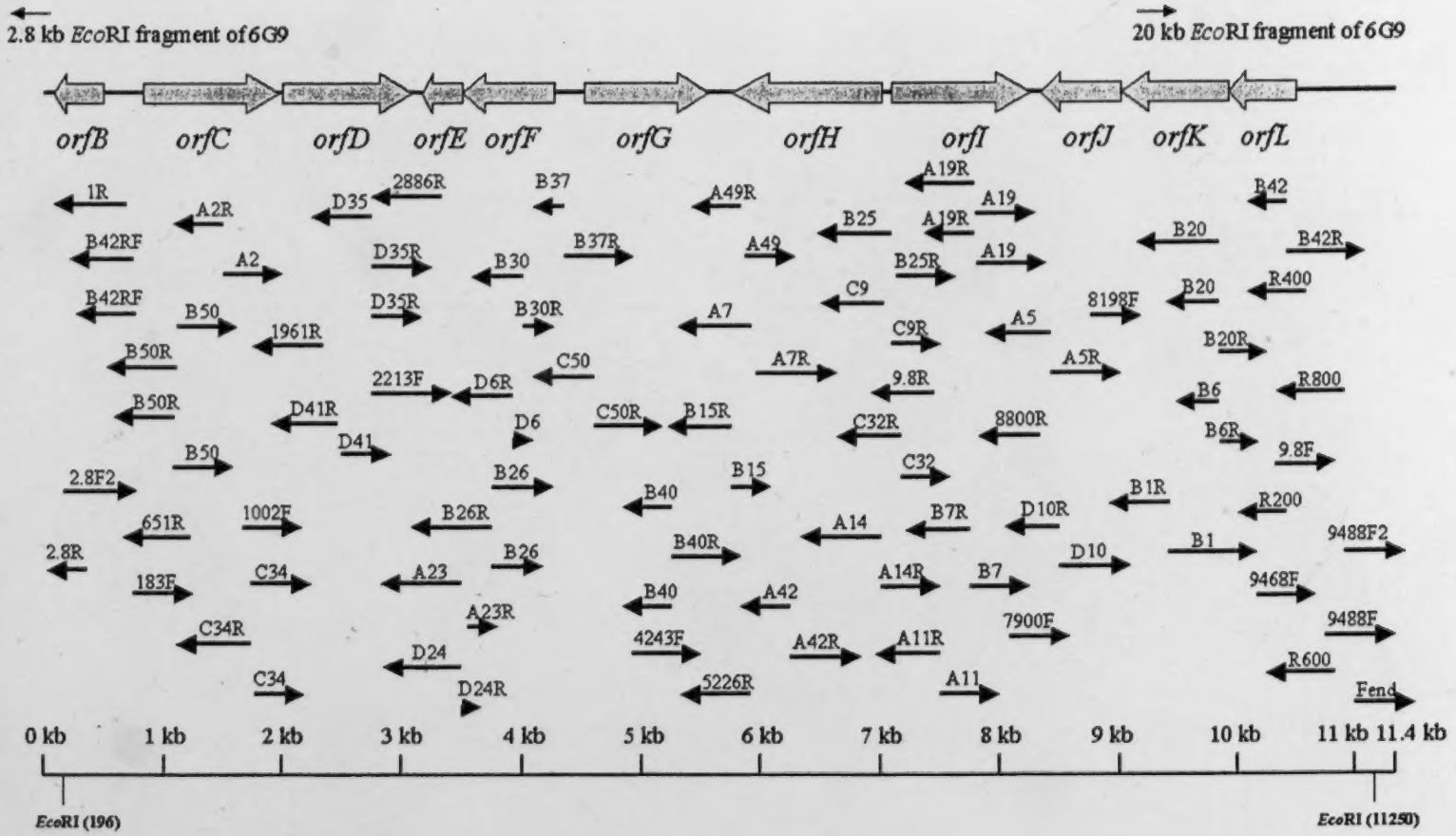
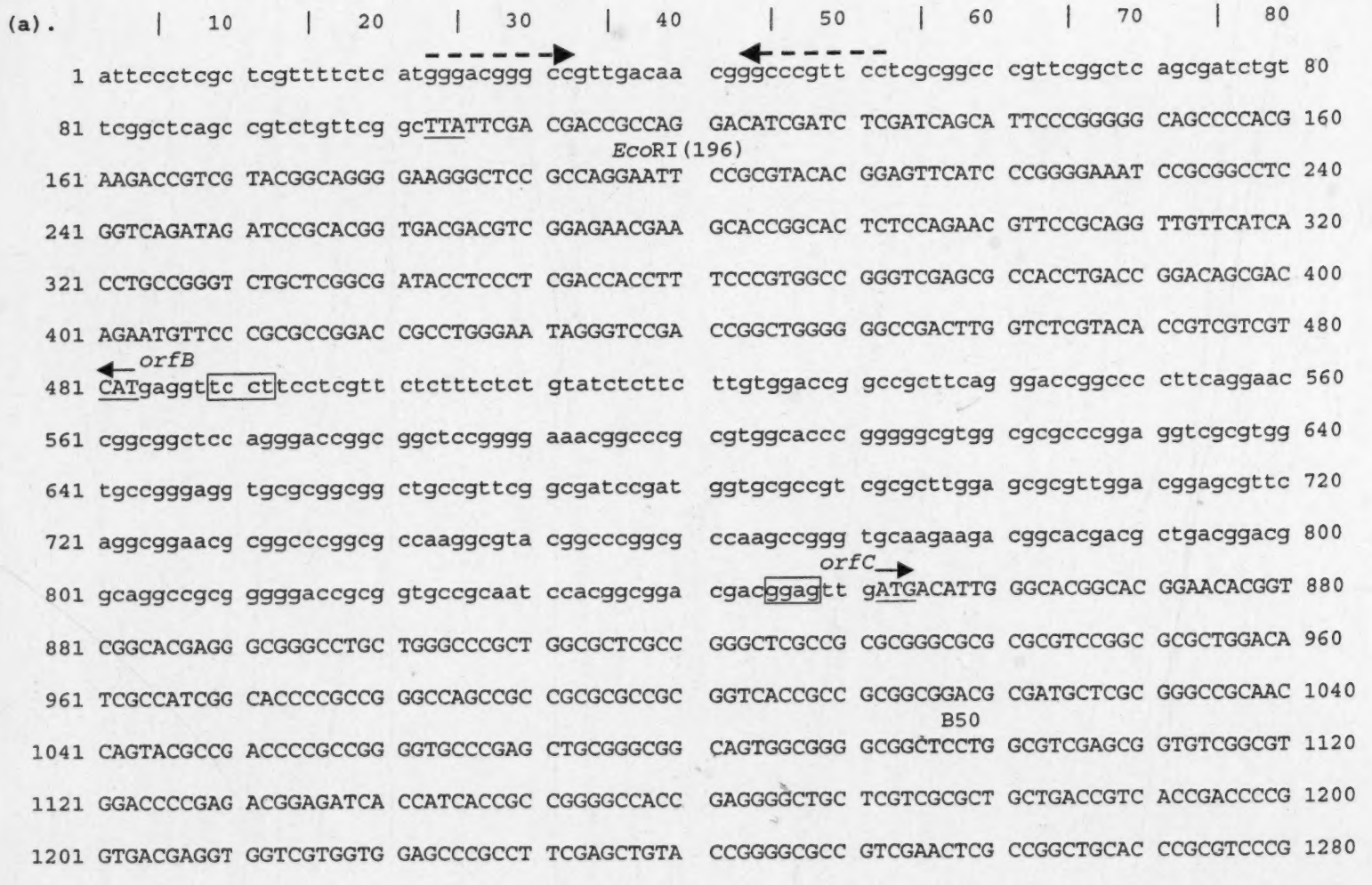
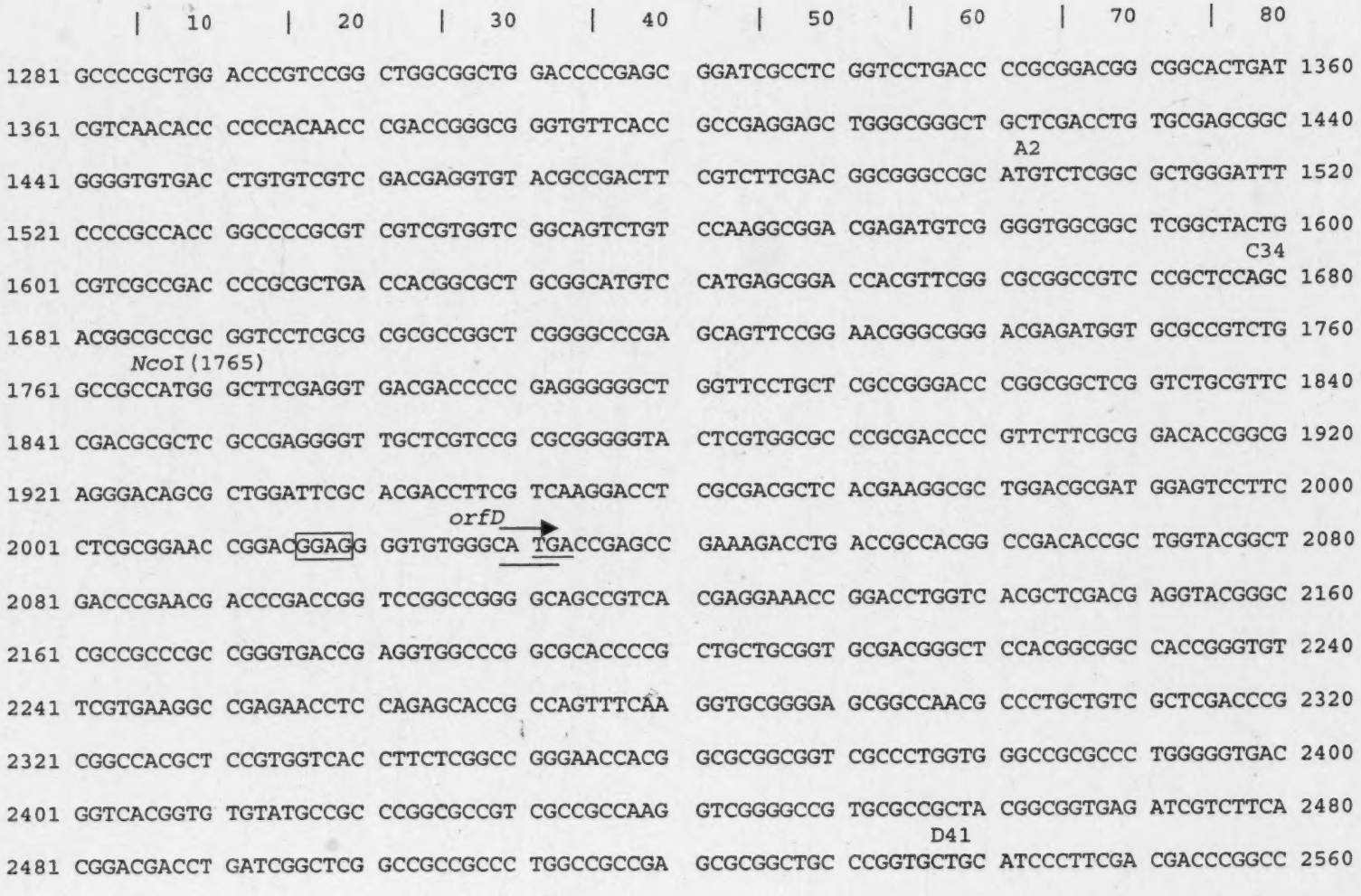
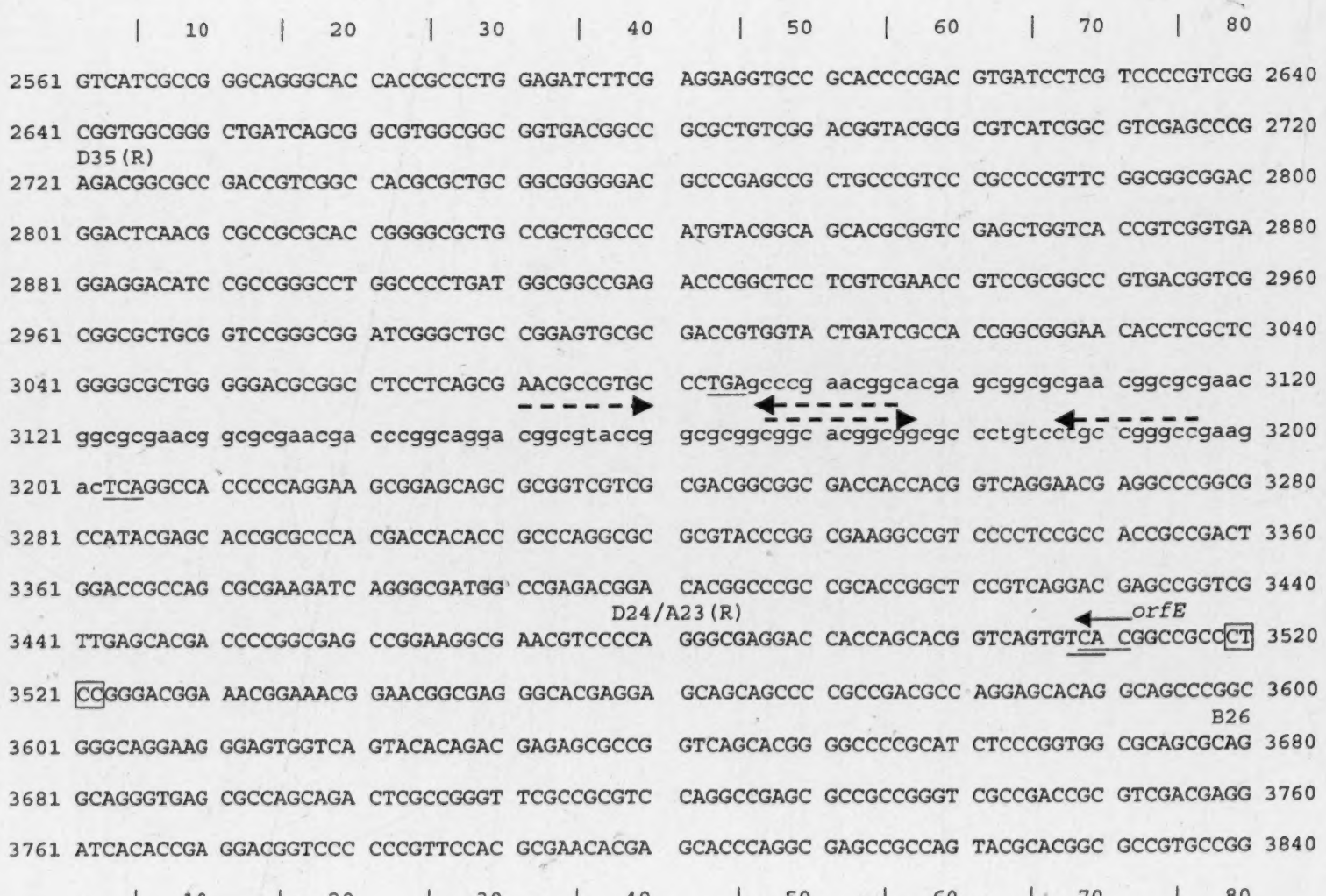
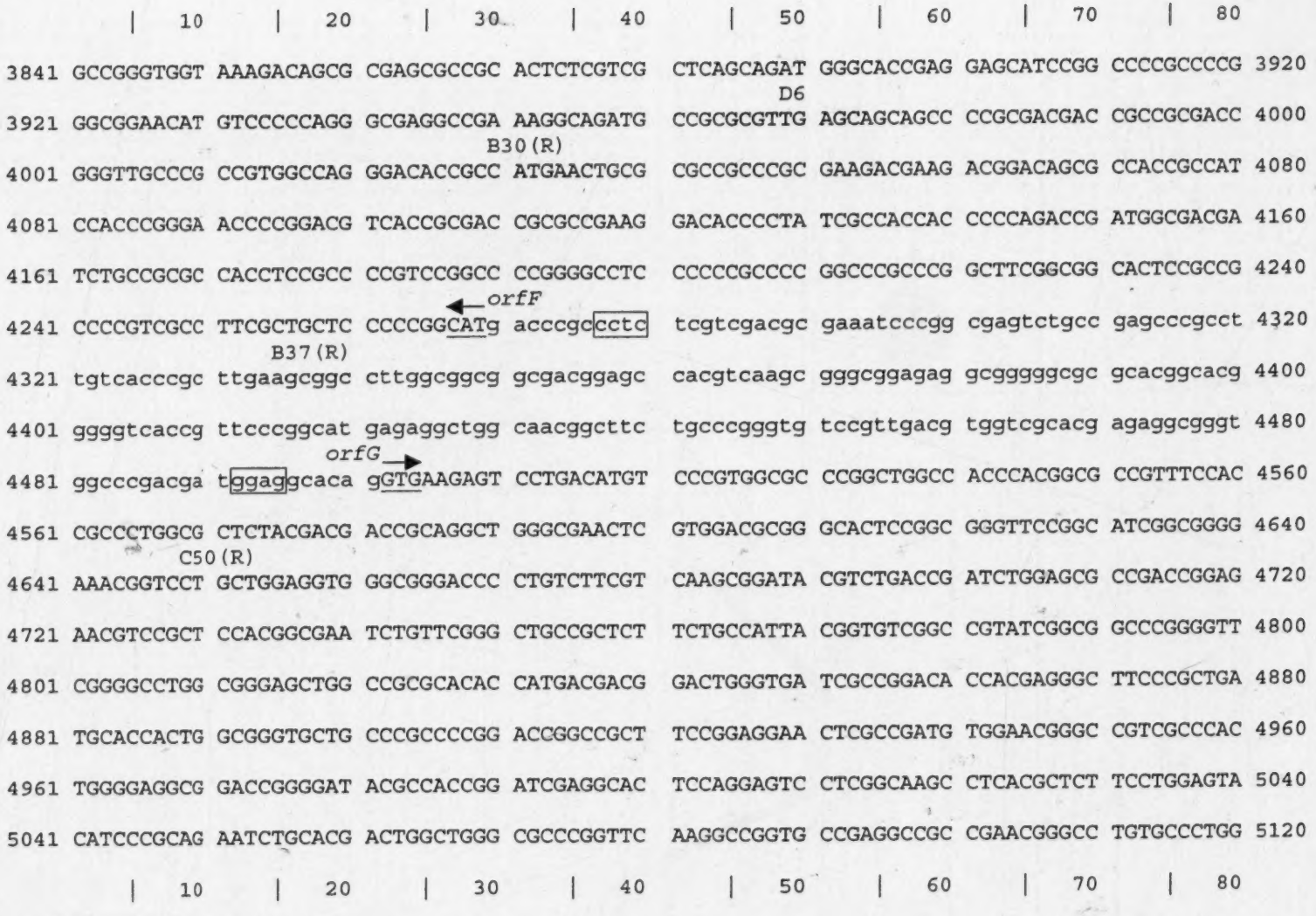

2.8 kb EcoRI fragment of 6G9
20 kb EcoRI fragment of 6G9
orfB
orfC
orfD
orfE
orfF
orfG
orfH
orfI
orfJ
orfK
orfL
1R
A2R
2886R
D35
B37
A49R
A19R
A19
B42
B25
A19R
B20
B42R
B42RF
A2
D35R
B37R
A49
B25R
A19
B30
C9
B20
R400
B42RF
D35R
8198F
B50
B30R
A7
A5
1961R
C9R
B20R
B50R
A7R
A5R
C50
9.8R
2213F
D6R
B6
R800
B50R
D41R
C50R
B15R
C32R
8800R
D6
B6R
D41
B50
9.8F
C32
2.8F2
B26
B15
B40
B1R
R200
651R
1002F
B26R
A14
B7R
D10R
9488F2
2.8R
B40R
D10
B1
B26
183F
C34
A23
A14R
B7
9468F
B40
A42
A23R
7900F
9488F
C34R
4243F
A42R
A11R
D24
R600
C34
5226R
A11
Fend
D24R
0 kb
1 kb
2 kb
3 kb
4 kb
5 kb
6 kb
7 kb
8 kb
9 kb
10 kb
11 kb
11.4 kb
EcoRI (196)
EcoRI (11250)

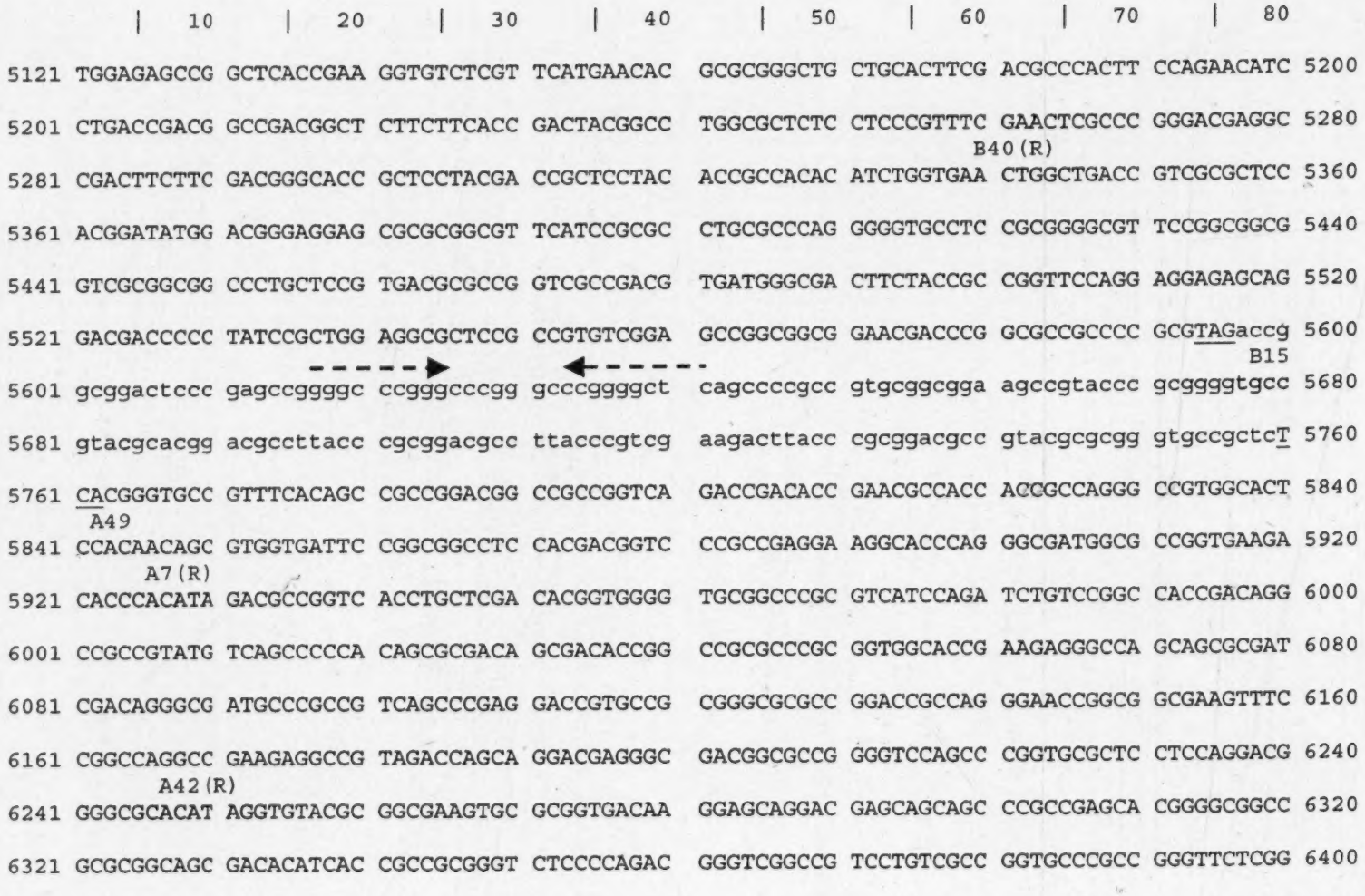
Figure 17. (a). Compiled 11.4 kb nucleotide sequence data. Regions of sequence containing ORFs are presented in upper case text while the intergenic regions are given in lower case text. The gene designations are labeled. The orientation of transcription is indicated by small arrows. Putative ribosome-binding sites are in open boxes. The predicted translational start codons and translational stop codons are underlined. Inverted repeats which may act as transcriptional terminators are indicated by dashed arrows. The *Eco*RI sites and *Nco*I sites together with the position of the cleavage sites are labeled. The transposon target sites are shadowed, and for those containing Tn*5* inserted into the opposite strand, a letter (R) was labeled with their designations. (b). Amino acid sequence for each identified *orfs*.

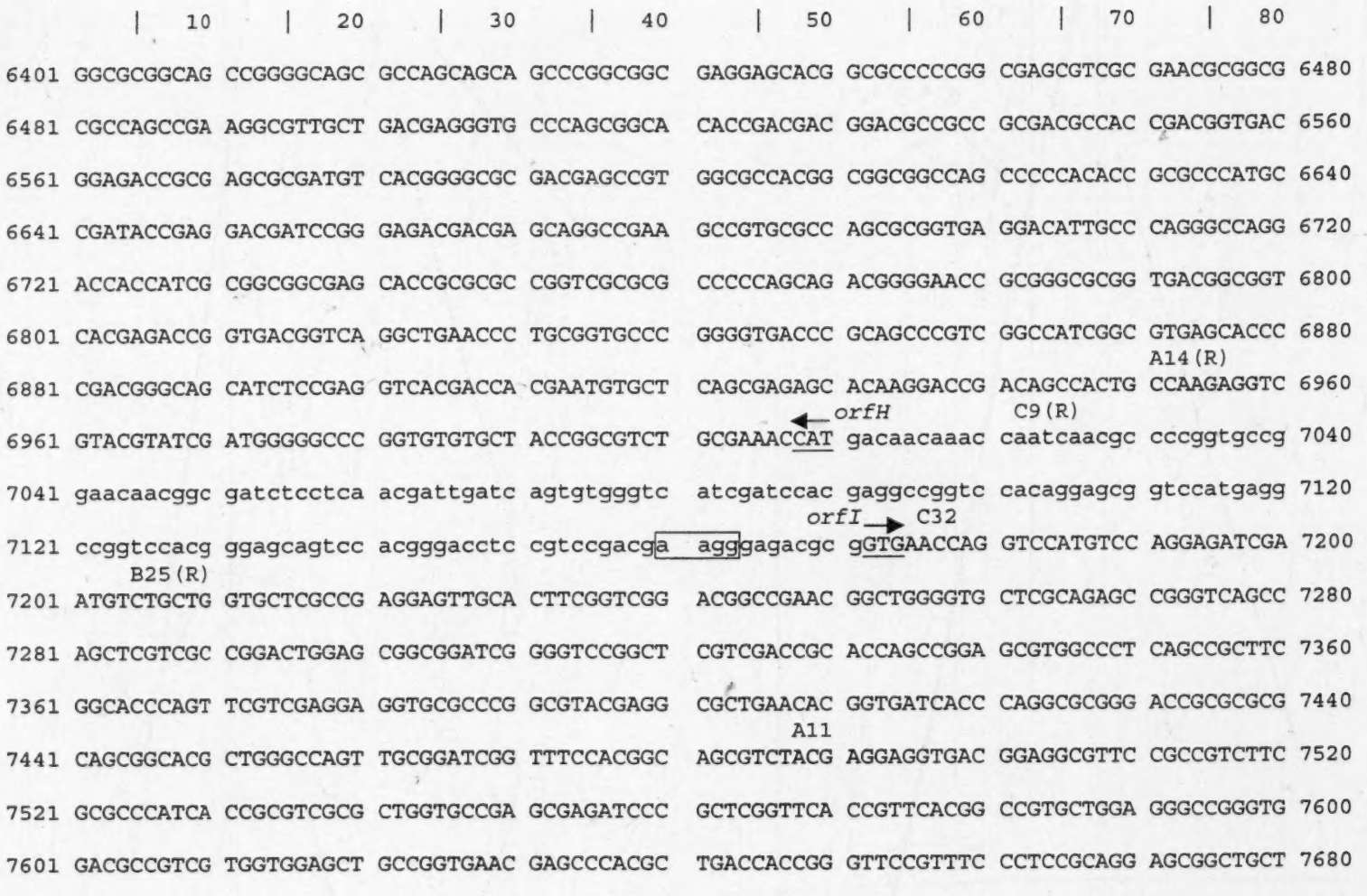

(a).
| 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80
1 attccctcgc tcgttttctc atgggacggg ccgttgacaa cgggcccgtt cctcgcggcc cgttcggctc agcgatctgt 80
81 tcggctcagc cgtctgttcg gcTTATTCGA CGACCGCCAG GACATCGATC TCGATCAGCA TTCCCGGGGG CAGCCCCACG 160
EcoRI(196)
161 AAGACCGTCG TACGGCAGGG GAAGGGCTCC GCCAGGAATT CCGCGTACAC GGAGTTCATC CCGGGGAAAT CCGCGGCCTC 240
241 GGTCAGATAG ATCCGCACGG TGACGACGTC GGAGAACGAA GCACCGGCAC TCTCCAGAAC GTTCCGCAGG TTGTTCATCA 320
321 CCTGCCGGGT CTGCTCGGCG ATACCTCCCT CGACCACCTT TCCCGTGGCC GGGTCGAGCG CCACCTGACC GGACAGCGAC 400
401 AGAATGTTCC CGCGCCGGAC CGCCTGGGAA TAGGGTCCGA CCGGCTGGGG GGCCGACTTG GTCTCGTACA CCGTCGTCGT 480
orfB
481 CATgaggttc ctcctcgtt ctctttctct gtatctcttc ttgtggaccg gccgcttcag ggaccggccc cttcaggaac 560
561 cggcggctcc agggaccggc ggctccgggg aaacggcccg cgtggcaccc gggggcgtgg cgcgcccgga ggtcgcgtgg 640
641 tgccgggagg tgcgcggcgg ctgccgttcg gcgatccgat ggtgcgccgt cgcgcttgga gcgcgttgga cggagcgttc 720
721 aggcggaacg cggcccggcg ccaaggcgta cggcccggcg ccaagccggg tgcaagaaga cggcacgacg ctgacggacg 800
orfC
801 gcaggccgcg ggggaccgcg gtgccgcaat ccacggcgga cgacggagtt gATGACATTG GGCACGGCAC GGAACACGGT 880
881 CGGCACGAGG GCGGGCCTGC TGGGCCCGCT GGCGCTCGCC GGGCTCGCCG CGCGGGCGCG CGCGTCCGGC GCGCTGGACA 960
961 TCGCCATCGG CACCCCGCCG GGCCAGCCGC CGCGCGCCGC GGTCACCGCC GCGGCGGACG CGATGCTCGC GGGCCGCAAC 1040
B50
1041 CAGTACGCCG ACCCCGCCGG GGTGCCCGAG CTGCGGGCGG CAGTGGCGGG GCGGCTCCTG GCGTCGAGCG GTGTCGGCGT 1120
1121 GGACCCCGAG ACGGAGATCA CCATCACCGC CGGGGCCACC GAGGGGCTGC TCGTCGCGCT GCTGACCGTC ACCGACCCCG 1200
1201 GTGACGAGGT GGTCGTGGTG GAGCCCGCCT TCGAGCTGTA CCGGGGCGCC GTCGAACTCG CCGGCTGCAC CCGCGTCCCG 1280
| 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80

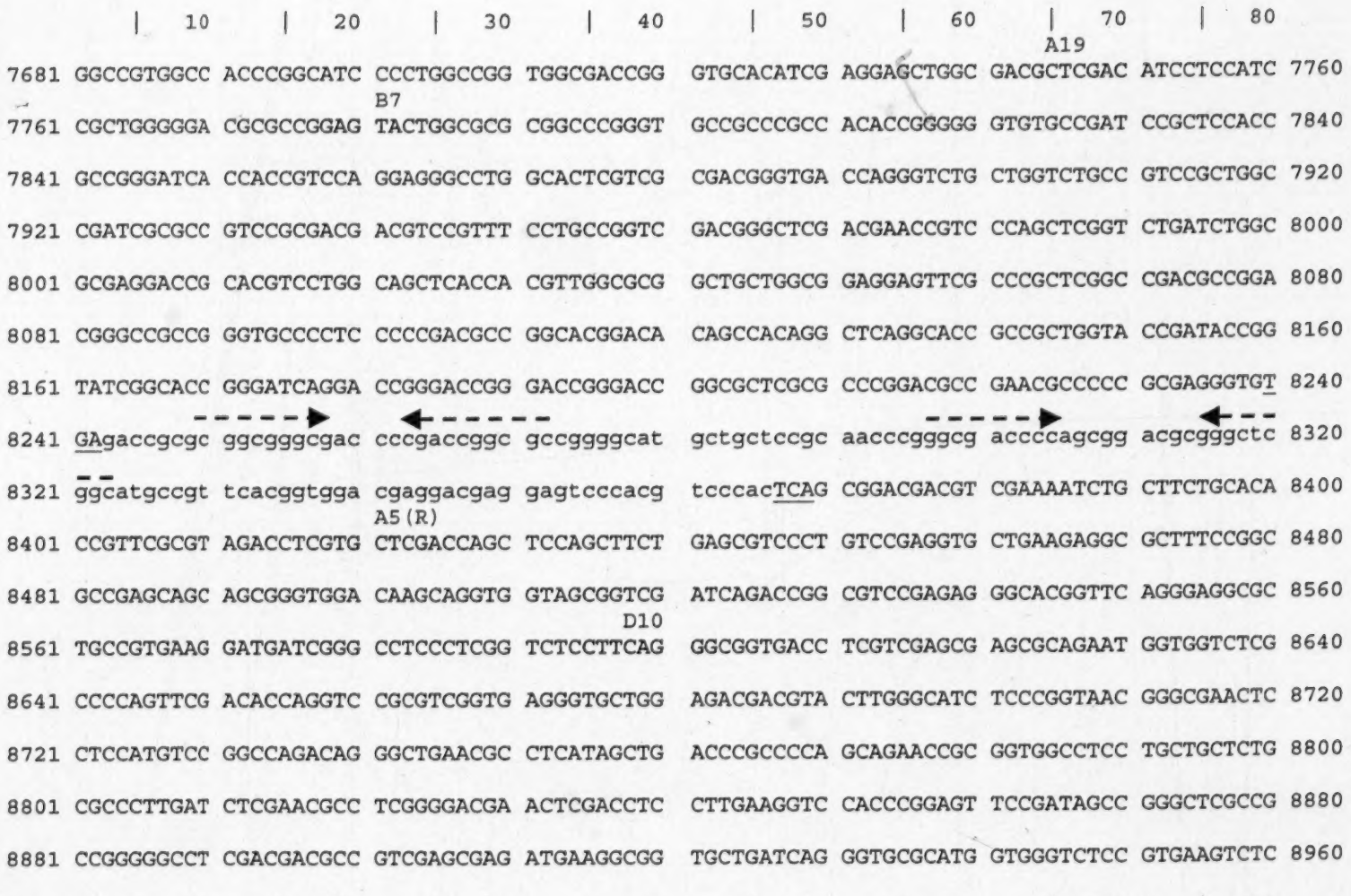

| 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80
1281 GCCCCGCTGG ACCCGTCCGG CTGGCGGCTG GACCCCGAGC GGATCGCCTC GGTCCTGACC CCGCGGACGG CGGCACTGAT 1360
1361 CGTCAACACC CCCCACAACC CGACCGGGCG GGTGTTCACC GCCGAGGAGC TGGGCGGGCT GCTCGACCTG TGCGAGCGGC 1440
A2
1441 GGGGTGTGAC CTGTGTCGTC GACGAGGTGT ACGCCGACTT CGTCTTCGAC GGCGGGCCGC ATGTCTCGGC GCTGGGATTT 1520
1521 CCCCGCCACC GGCCCCGCGT CGTCGTGGTC GGCAGTCTGT CCAAGGCGGA CGAGATGTCG GGGTGGCGGC TCGGCTACTG 1600
C34
1601 CGTCGCCGAC CCCGCGCTGA CCACGGCGCT GCGGCATGTC CATGAGCGGA CCACGTTCGG CGCGGCCGTC CCGCTCCAGC 1680
1681 ACGGCGCCGC GGTCCTCGCG CGCGCCGGCT CGGGGCCCGA GCAGTTCCGG AACGGGCGGG ACGAGATGGT GCGCCGTCTG 1760
NcoI(1765)
1761 GCCGCCATGG GCTTCGAGGT GACGACCCCC GAGGGGGGCT GGTTCCTGCT CGCCGGGACC CGGCGGCTCG GTCTGCGTTC 1840
1841 CGACGCGCTC GCCGAGGGGT TGCTCGTCCG CGCGGGGGTA CTCGTGGCGC CCGCGACCCC GTTCTTCGCG GACACCGGCG 1920
1921 AGGGACAGCG CTGGATTCGC ACGACCTTCG TCAAGGACCT CGCGACGCTC ACGAAGGCGC TGGACGCGAT GGAGTCCTTC 2000
orfD
2001 CTCGCGGAAC CGGACGGAGG GGTGTGGGCA TGACCGAGCC GAAAGACCTG ACCGCCACGG CCGACACCGC TGGTACGGCT 2080
2081 GACCCGAACG ACCCGACCGG TCCGGCCGGG GCAGCCGTCA CGAGGAAACC GGACCTGGTC ACGCTCGACG AGGTACGGGC 2160
2161 CGCCGCCCGC CGGGTGACCG AGGTGGCCCG GCGCACCCCG CTGCTGCGGT GCGACGGGCT CCACGGCGGC CACCGGGTGT 2240
2241 TCGTGAAGGC CGAGAACCTC CAGAGCACCG CCAGTTTCAA GGTGCGGGGA GCGGCCAACG CCCTGCTGTC GCTCGACCCG 2320
2321 CGGCCACGCT CCGTGGTCAC CTTCTCGGCC GGGAACCACG GCGCGGCGGT CGCCCTGGTG GGCCGCGCCC TGGGGGTGAC 2400
2401 GGTCACGGTG TGTATGCCGC CCGGCGCCGT CGCCGCCAAG GTCGGGGCCG TGCGCCGCTA CGGCGGTGAG ATCGTCTTCA 2480
D41
2481 CGGACGACCT GATCGGCTCG GCCGCCGCCC TGGCCGCCGA GCGCGGCTGC CCGGTGCTGC ATCCCTTCGA CGACCCGGCC 2560
| 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80

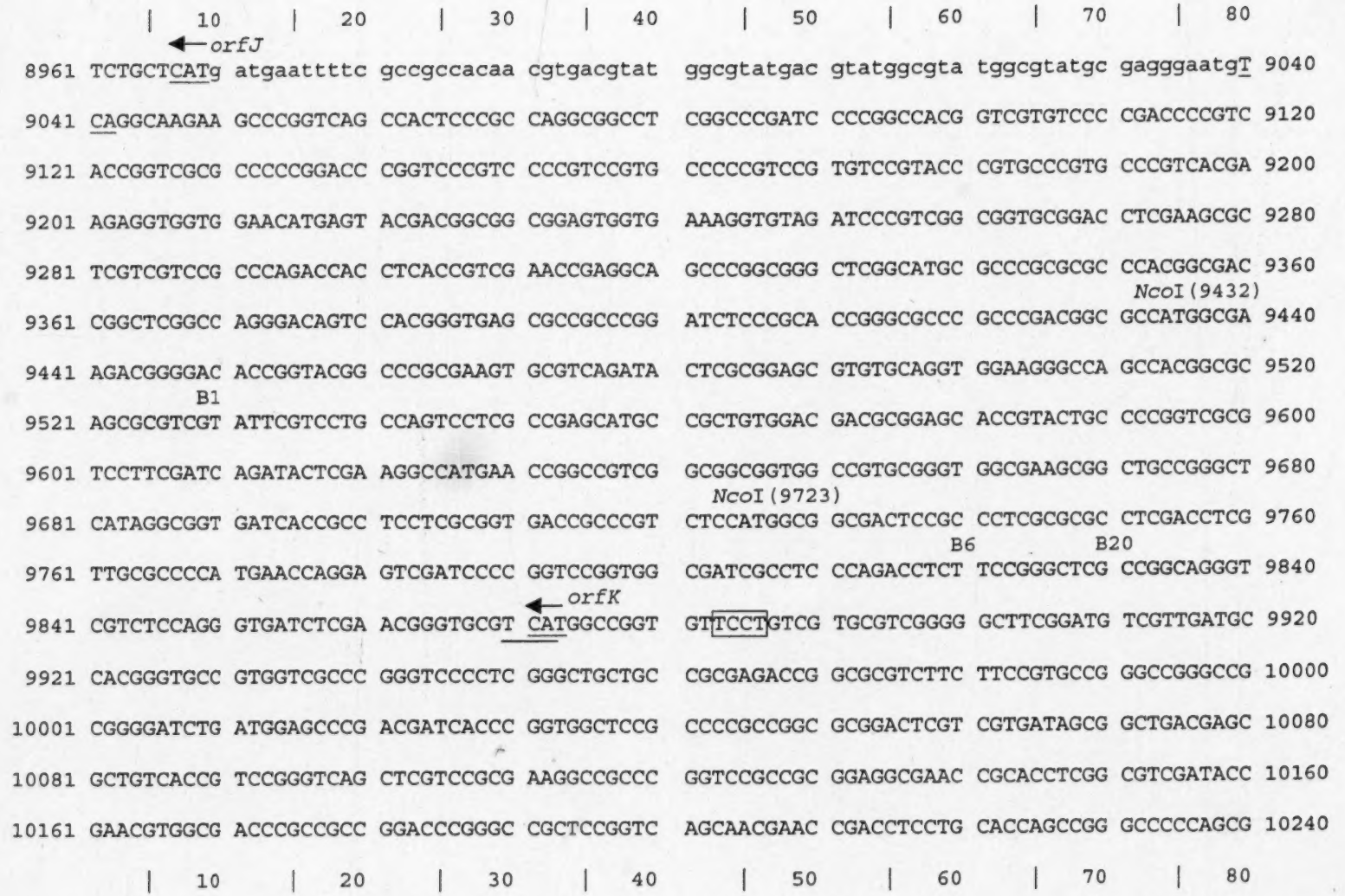

| 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80
2561 GTCATCGCCG GGCAGGGCAC CACCGCCCTG GAGATCTTCG AGGAGGTGCC GCACCCCGAC GTGATCCTCG TCCCCGTCGG 2640
2641 CGGTGGCGGG CTGATCAGCG GCGTGGCGGC GGTGACGGCC GCGCTGTCGG ACGGTACGCG CGTCATCGGC GTCGAGCCCG 2720
D35(R)
2721 AGACGGCGCC GACCGTCGGC CACGCGCTGC GGCGGGGGAC GCCCGAGCCG CTGCCCGTCC CGCCCCGTTC GGCGGCGGAC 2800
2801 GGACTCAACG CGCCGCGCAC CGGGGCGCTG CCGCTCGCCC ATGTACGGCA GCACGCGGTC GAGCTGGTCA CCGTCGGTGA 2880
2881 GGAGGACATC CGCCGGGCCT GGCCCCTGAT GGCGGCCGAG ACCCGGCTCC TCGTCGAACC GTCCGCGGCC GTGACGGTCG 2960
2961 CGGCGCTGCG GTCCGGGCGG ATCGGGCTGC CGGAGTGCGC GACCGTGGTA CTGATCGCCA CCGGCGGGAA CACCTCGCTC 3040
3041 GGGGCGCTGG GGGACGCGGC CTCCTCAGCG AACGCCGTGC CCTGAgcccg aacggcacga gcggcgcgaa cggcgcgaac 3120
3121 ggcgcgaacg gcgcgaacga cccggcagga cggcgtaccg gcgcggcggc acggcggcgc cctgtcctgc cgggccgaag 3200
3201 acTCAGGCCA CCCCCAGGAA GCGGAGCAGC GCGGTCGTCG CGACGGCGGC GACCACCACG GTCAGGAACG AGGCCCGGCG 3280
3281 CCATACGAGC ACCGCGCCCA CGACCACACC GCCCAGGCGC GCGTACCCGG CGAAGGCCGT CCCCTCCGCC ACCGCCGACT 3360
3361 GGACCGCCAG CGCGAAGATC AGGGCGATGG CCGAGACGGA CACGGCCCGC CGCACCGGCT CCGTCAGGAC GAGCCGGTCG 3440
D24/A23(R)
orfE
3441 TTGAGCACGA CCCCGGCGAG CCGGAAGGCG AACGTCCCCA GGGCGAGGAC CACCAGCACG GTCAGTGTCA CGGCCGCCCT 3520
3521 CCGGGACGGA AACGGAAACG GAACGGCGAG GGCACGAGGA GCAGCAGCCC CGCCGACGCC AGGAGCACAG GCAGCCCGGC 3600
B26
3601 GGGCAGGAAG GGAGTGGTCA GTACACAGAC GAGAGCGCCG GTCAGCACGG GGCCCCGCAT CTCCCGGTGG CGCAGCGCAG 3680
3681 GCAGGGTGAG CGCCAGCAGA CTCGCCGGGT TCGCCGCGTC CAGGCCGAGC GCCGCCGGGT CGCCGACCGC GTCGACGAGG 3760
3761 ATCACACCGA GGACGGTCCC CCCGTTCCAC GCGAACACGA GCACCCAGGC GAGCCGCCAG TACGCACGGC GCCGTGCCGG 3840
| 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80


| 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80
3841 GCCGGGTGGT AAAGACAGCG CGAGCGCCGC ACTCTCGTCG CTCAGCAGAT GGGCACCGAG GAGCATCCGG CCCCGCCCCG 3920
D6
3921 GGCGGAACAT GTCCCCCAGG GCGAGGCCGA AAGGCAGATG CCGCGCGTTG AGCAGCAGCC CCGCGACGAC CGCCGCGACC 4000
B30(R)
4001 GGGTTGCCCG CCGTGGCCAG GGACACCGCC ATGAACTGCG CGCCGCCCGC GAAGACGAAG ACGGACAGCG CCACCGCCAT 4080
4081 CCACCCGGGA ACCCCGGACG TCACCGCGAC CGCGCCGAAG GACACCCCTA TCGCCACCAC CCCCAGACCG ATGGCGACGA 4160
4161 TCTGCCGCGC CACCTCCGCC CCGTCCGGCC CCGGGGCCTC CCCCCGCCCC GGCCCGCCCG GCTTCGGCGG CACTCCGCCG 4240
orfF
4241 CCCCGTCGCC TTCGCTGCTC CCCCGGCATg acccgccctc tcgtcgacgc gaaatcccgg cgagtctgcc gagcccgcct 4320
B37(R)
4321 tgtcacccgc ttgaagcggc cttggcggcg gcgacggagc cacgtcaagc gggcggagag gcgggggcgc gcacggcacg 4400
4401 ggggtcaccg ttcccggcat gagaggctgg caacggcttc tgcccgggtg tccgttgacg tggtcgcacg agaggcgggt 4480
orfG
4481 ggcccgacga tggaggcaca gGTGAAGAGT CCTGACATGT CCCGTGGCGC CCGGCTGGCC ACCCACGGCG CCGTTTCCAC 4560
4561 CGCCCTGGCG CTCTACGACG ACCGCAGGCT GGGCGAACTC GTGGACGCGG GCACTCCGGC GGGTTCCGGC ATCGGCGGGG 4640
C50(R)
4641 AAACGGTCCT GCTGGAGGTG GGCGGGACCC CTGTCTTCGT CAAGCGGATA CGTCTGACCG ATCTGGAGCG CCGACCGGAG 4720
4721 AACGTCCGCT CCACGGCGAA TCTGTTCGGG CTGCCGCTCT TCTGCCATTA CGGTGTCGGC CGTATCGGCG GCCCGGGGTT 4800
4801 CGGGGCCTGG CGGGAGCTGG CCGCGCACAC CATGACGACG GACTGGGTGA TCGCCGGACA CCACGAGGGC TTCCCGCTGA 4880
4881 TGCACCACTG GCGGGTGCTG CCCGCCCCGG ACCGGCCGCT TCCGGAGGAA CTCGCCGATG TGGAACGGGC CGTCGCCCAC 4960
4961 TGGGGAGGCG GACCGGGGAT ACGCCACCGG ATCGAGGCAC TCCAGGAGTC CTCGGCAAGC CTCACGCTCT TCCTGGAGTA 5040
5041 CATCCCGCAG AATCTGCACG ACTGGCTGGG CGCCCGGTTC AAGGCCGGTG CCGAGGCCGC CGAACGGGCC TGTGCCCTGG 5120
| 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80


| 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80
5121 TGGAGAGCCG GCTCACCGAA GGTGTCTCGT TCATGAACAC GCGCGGGCTG CTGCACTTCG ACGCCCACTT CCAGAACATC 5200
5201 CTGACCGACG GCCGACGGCT CTTCTTCACC GACTACGGCC TGGCGCTCTC CTCCCGTTTC GAACTCGCCC GGGACGAGGC 5280
B40(R)
5281 CGACTTCTTC GACGGGCACC GCTCCTACGA CCGCTCCTAC ACCGCCACAC ATCTGGTGAA CTGGCTGACC GTCGCGCTCC 5360
5361 ACGGATATGG ACGGGAGGAG CGCGCGGCGT TCATCCGCGC CTGCGCCCAG GGGGTGCCTC CGCGGGGCGT TCCGGCGGCG 5440
5441 GTCGCGGCGG CCCTGCTCCG TGACGCGCCG GTCGCCGACG TGATGGGCGA CTTCTACCGC CGGTTCCAGG AGGAGAGCAG 5520
5521 GACGACCCCC TATCCGCTGG AGGCGCTCCG CCGTGTCGGA GCCGGCGGCG GAACGACCCG GCGCCGCCCC GCGTAGaccg 5600
B15
5601 gcggactccc gagccggggc ccgggcccgg gcccggggct cagccccgcc gtgcggcgga agccgtaccc gcggggtgcc 5680
5681 gtacgcacgg acgccttacc cgcggacgcc ttacccgtcg aagacttacc cgcggacgcc gtacgcgcgg gtgccgctcT 5760
5761 CACGGGTGCC GTTTCACAGC CGCCGGACGG CCGCCGGTCA GACCGACACC GAACGCCACC ACCGCCAGGG CCGTGGCACT 5840
A49
5841 CCACAACAGC GTGGTGATTC CGGCGGCCTC CACGACGGTC CCGCCGAGGA AGGCACCCAG GGCGATGGCG CCGGTGAAGA 5920
A7(R)
5921 CACCCACATA GACGCCGGTC ACCTGCTCGA CACGGTGGGG TGCGGCCCGC GTCATCCAGA TCTGTCCGGC CACCGACAGG 6000
6001 CCGCCGTATG TCAGCCCCCA CAGCGCGACA GCGACACCGG CCGCGCCCGC GGTGGCACCG AAGAGGGCCA GCAGCGCGAT 6080
6081 CGACAGGGCG ATGCCCGCCG TCAGCCCGAG GACCGTGCCG CGGGCGCGCC GGACCGCCAG GGAACCGGCG GCGAAGTTTC 6160
6161 CGGCCAGGCC GAAGAGGCCG TAGACCAGCA GGACGAGGGC GACGGCGCCG GGGTCCAGCC CGGTGCGCTC CTCCAGGACG 6240
A42(R)
6241 GGGCGCACAT AGGTGTACGC GGCGAAGTGC GCGGTGACAA GGAGCAGGAC GAGCAGCAGC CCGCCGAGCA CGGGGCGGCC 6320
6321 GCGCGGCAGC GACACATCAC CGCCGCGGGT CTCCCCAGAC GGGTCGGCCG TCCTGTCGCC GGTGCCCGCC GGGTTCTCGG 6400
| 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80


| 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80
6401 GGCGCGGCAG CCGGGGCAGC GCCAGCAGCA GCCCGGCGGC GAGGAGCACG GCGCCCCCGG CGAGCGTCGC GAACGCGGCG 6480
6481 CGCCAGCCGA AGGCGTTGCT GACGAGGGTG CCCAGCGGCA CACCGACGAC GGACGCCGCC GCGACGCCAC CGACGGTGAC 6560
6561 GGAGACCGCG AGCGCGATGT CACGGGGCGC GACGAGCCGT GGCGCCACGG CGGCGGCCAG CCCCCACACC GCGCCCATGC 6640
6641 CGATACCGAG GACGATCCGG GAGACGACGA GCAGGCCGAA GCCGTGCGCC AGCGCGGTGA GGACATTGCC CAGGGCCAGG 6720
6721 ACCACCATCG CGGCGGCGAG CACCGCGCGC CGGTCGCGCG CCCCCAGCAG ACGGGGAACC GCGGGCGCGG TGACGGCGGT 6800
6801 CACGAGACCG GTGACGGTCA GGCTGAACCC TGCGGTGCCC GGGGTGACCC GCAGCCCGTC GGCCATCGGC GTGAGCACCC 6880
A14(R)
6881 CGACGGGCAG CATCTCCGAG GTCACGACCA CGAATGTGCT CAGCGAGAGC ACAAGGACCG ACAGCCACTG CCAAGAGGTC 6960
orfH
C9(R)
6961 GTACGTATCG ATGGGGGCCC GGTGTGTGCT ACCGGCGTCT GCGAAACCAT gacaacaaac caatcaacgc cccggtgccg 7040
7041 gaacaacggc gatctcctca acgattgatc agtgtgggtc atcgatccac gaggccggtc cacaggagcg gtccatgagg 7120
orfI
C32
7121 ccggtccacg ggagcagtcc acgggacctc cgtccgacga aggggagacgc gGTGAACCAG GTCCATGTCC AGGAGATCGA 7200
B25(R)
7201 ATGTCTGCTG GTGCTCGCCG AGGAGTTGCA CTTCGGTCGG ACGGCCGAAC GGCTGGGGTG CTCGCAGAGC CGGGTCAGCC 7280
7281 AGCTCGTCGC CGGACTGGAG CGGCGGATCG GGGTCCGGCT CGTCGACCGC ACCAGCCGGA GCGTGGCCCT CAGCCGCTTC 7360
7361 GGCACCCAGT TCGTCGAGGA GGTGCGCCCG GCGTACGAGG CGCTGAACAC GGTGATCACC CAGGCGCGGG ACCGCGCGCG 7440
A11
7441 CAGCGGCACG CTGGGCCAGT TGCGGATCGG TTTCCACGGC AGCGTCTACG AGGAGGTGAC GGAGGCGTTC CGCCGTCTTC 7520
7521 GCGCCCATCA CCGCGTCGCG CTGGTGCCGA GCGAGATCCC GCTCGGTTCA CCGTTCACGG CCGTGCTGGA GGGCCGGGTG 7600
7601 GACGCCGTCG TGGTGGAGCT GCCGGTGAAC GAGCCCACGC TGACCACCGG GTTCCGTTTC CCTCCGCAGG AGCGGCTGCT 7680
| 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80


| 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80
A19
7681 GGCCGTGGCC ACCCGGCATC CCCTGGCCGG TGGCGACCGG GTGCACATCG AGGAGCTGGC GACGCTCGAC ATCCTCCATC 7760
B7
7761 CGCTGGGGGA CGCGCCGGAG TACTGGCGCG CGGCCCGGGT GCCGCCCGCC ACACCGGGGG GTGTGCCGAT CCGCTCCACC 7840
7841 GCCGGGATCA CCACCGTCCA GGAGGGCCTG GCACTCGTCG CGACGGGTGA CCAGGGTCTG CTGGTCTGCC GTCCGCTGGC 7920
7921 CGATCGCGCC GTCCGCGACG ACGTCCGTTT CCTGCCGGTC GACGGGCTCG ACGAACCGTC CCAGCTCGGT CTGATCTGGC 8000
8001 GCGAGGACCG CACGTCCTGG CAGCTCACCA CGTTGGCGCG GCTGCTGGCG GAGGAGTTCG CCCGCTCGGC CGACGCCGGA 8080
8081 CGGGCCGCCG GGTGCCCCTC CCCCGACGCC GGCACGGACA CAGCCACAGG CTCAGGCACC GCCGCTGGTA CCGATACCGG 8160
8161 TATCGGCACC GGGATCAGGA CCGGGACCGG GACCGGGACC GGCGCTCGCG CCCGGACGCC GAACGCCCCC GCGAGGGTGT 8240
8241 GAgaccgcgc ggcgggcgac cccgaccggc gccggggcat gctgctccgc aacccgggcg accccagcgg acgcgggctc 8320
8321 ggcatgccgt tcacggtgga cgaggacgag gagtcccacg tcccacTCAG CGGACGACGT CGAAAATCTG CTTCTGCACA 8400
A5(R)
8401 CCGTTCGCGT AGACCTCGTG CTCGACCAGC TCCAGCTTCT GAGCGTCCCT GTCCGAGGTG CTGAAGAGGC GCTTTCCGGC 8480
8481 GCCGAGCAGC AGCGGGTGGA CAAGCAGGTG GTAGCGGTCG ATCAGACCGG CGTCCGAGAG GGCACGGTTC AGGGAGGCGC 8560
D10
8561 TGCCGTGAAG GATGATCGGG CCTCCCTCGG TCTCCTTCAG GGCGGTGACC TCGTCGAGCG AGCGCAGAAT GGTGGTCTCG 8640
8641 CCCCAGTTCG ACACCAGGTC CGCGTCGGTG AGGGTGCTGG AGACGACGTA CTTGGGCATC TCCCGGTAAC GGGCGAACTC 8720
8721 CTCCATGTCC GGCCAGACAG GGCTGAACGC CTCATAGCTG ACCCGCCCCA GCAGAACCGC GGTGGCCTCC TGCTGCTCTG 8800
8801 CGCCCTTGAT CTCGAACGCC TCGGGGACGA ACTCGACCTC CTTGAAGGTC CACCCGGAGT TCCGATAGCC GGGCTCGCCG 8880
8881 CCGGGGGCCT CGACGACGCC GTCGAGCGAG ATGAAGGCGG TGCTGATCAG GGTGCGCATG GTGGGTCTCC GTGAAGTCTC 8960
| 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80


10 20 30 40 50 60 70 80
orfJ
8961 TCTGCTCATg atgaattttc gccgccacaa cgtgacgtat ggcgtatgac gtatggcgta tggcgtatgc gagggaatgT 9040
9041 CAGGCAAGAA GCCCGGTCAG CCACTCCCGC CAGGCGGCCT CGGCCCGATC CCCGGCCACG GTCGTGTCCC CGACCCCGTC 9120
9121 ACCGGTCGCG CCCCCGGACC CGGTCCCGTC CCCGTCCGTG CCCCCGTCCG TGTCCGTACC CGTGCCCGTG CCCGTCACGA 9200
9201 AGAGGTGGTG GAACATGAGT ACGACGGCGG CGGAGTGGTG AAAGGTGTAG ATCCCGTCGG CGGTGCGGAC CTCGAAGCGC 9280
9281 TCGTCGTCCG CCCAGACCAC CTCACCGTCG AACCGAGGCA GCCCGGCGGG CTCGGCATGC GCCCGCGCGC CCACGGCGAC 9360
NcoI (9432)
9361 CGGCTCGGCC AGGGACAGTC CACGGGTGAG CGCCGCCCGG ATCTCCCGCA CCGGGCGCCC GCCCGACGGC GCCATGGCGA 9440
9441 AGACGGGGAC ACCGGTACGG CCCGCGAAGT GCGTCAGATA CTCGCGGAGC GTGTGCAGGT GGAAGGGCCA GCCACGGCGC 9520
B1
9521 AGCGCGTCGT ATTCGTCCTG CCAGTCCTCG CCGAGCATGC CGCTGTGGAC GACGCGGAGC ACCGTACTGC CCCGGTCGCG 9600
9601 TCCTTCGATC AGATACTCGA AGGCCATGAA CCGGCCGTCG GCGGCGGTGG CCGTGCGGGT GGCGAAGCGG CTGCCGGGCT 9680
NcoI (9723)
9681 CATAGGCGGT GATCACCGCC TCCTCGCGGT GACCGCCCGT CTCCATGGCG GCGACTCCGC CCTCGCGCGC CTCGACCTCG 9760
B6 B20
9761 TTGCGCCCCA TGAACCAGGA GTCGATCCCC GGTCCGGTGG CGATCGCCTC CCAGACCTCT TCCGGGCTCG CCGGCAGGGT 9840
orfK
9841 CGTCTCCAGG GTGATCTCGA ACGGGTGCGT CATGGCCGGT GTTCCTGTCG TGCGTCGGGG GCTTCGGATG TCGTTGATGC 9920
9921 CACGGGTGCC GTGGTCGCCC GGGTCCCCTC GGGCTGCTGC CGCGAGACCG GCGCGTCTTC TTCCGTGCCG GGCCGGGCCG 10000
10001 CGGGGATCTG ATGGAGCCCG ACGATCACCC GGTGGCTCCG CCCCGCCGGC GCGGACTCGT CGTGATAGCG GCTGACGAGC 10080
10081 GCTGTCACCG TCCGGGTCAG CTCGTCCGCG AAGGCCGCCC GGTCCGCCGC GGAGGCGAAC CGCACCTCGG CGTCGATACC 10160
10161 GAACGTGGCG ACCCGCCGCC GGACCCGGGC CGCTCCGGTC AGCAACGAAC CGACCTCCTG CACCAGCCGG GCCCCCAGCG 10240
10 20 30 40 50 60 70 80

```
          |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
10241 CCAGCAGCCA GCGGGCCGAG AGCTGGTCCG GGGAGCGCGA  GGGGTCCGGG CTGACAGCGG CCAGGGCGCT CGGCGAGATC 10320
                                                                                      B42
10321 ACATAGGAGG CGGCGGTCGC CCCGTACACC CGCTCGGTGA  CATTGCCCTT GCGCCGCTCC TCGACCAGCT CGACCAGACC 10400
10401 GTGCCGCTCC AGCTCCTTCA GGTGGTAGTT GACCTTCTGT  CTCGCCAGCC CCAGACGCGT GGCCAGCATG GCAGCCGAGC 10480
                                                                                  ←orfL
10481 CGGGCTCGGC CAGGGCGGCG AGGATCCGGG ACCGTATCGG  GTCCAGCGAA GCCTCGGCCG CGGCGGGTTC TTCGATCACa 10560
10561 gctacgtcca gcacacctcc agcatgggtc ggacaaaaaa  atttgtcaag gaaaaatgaa ctgtcggctc tgggctcgat 10640
10641 ccgctccact gggatcgtcg tcgcctccag tggcccctgc  cgcagatggg cctttgaccg gcgcagggcg ggccgggagg 10720
10721 cgcagcgcgg cgacggcgcc cgccgaggtg atcagagcgc  cgttgtcgat cgagagactg ccgcggtcga cctcgacggc 10800
10801 ggcgtgacgg cgggccggtg cggccgcgtg gtgcccgtga  gcggtggccc gctgcccgtc gggtcgaggc cctgcccggg 10880
                          NcoI(10906)
10881 gccgggatca gatatcagcc gtcagccatg ggctgccagg  tctttcacgg tctggagggt cggggcttcc tcgacggcgg 10960
10961 tcgaggcgat atgactcggc aggcgggggt cggagttcct  gagcagttcg gccagggctt ctgtcggcgc gacgccccgg 11040
11041 ctcctcgcgt cgcggagcgc ggcgacccac ctctgcccca  gattgaacgc atgccctccg cctgagagtc ggcggcacca 11120
11121 gcctgcggca tcgtcttcat cgttgcccca gatgtagccg  atggtctctc cgtccgcgtc ggccaccgta aggtatttca 11200
                                                        EcoRI(11251)
11201 ccgttttgtc cgtgctgttg gtgtaacgca gaggaccaga  gaccgacctg aattcgagat cctcagcaaa ccgaggagag 11280
11281 gtacgacgct cgcttccttc ggacactatt ctccctcacc  cttcacaact ggatttacga tttcagcgta tatgaaatac 11360
11361 cgaggccctg cgaaccctga tccatgacaa cactctcgac  ctggtatttc agaccgcgcg ccagcagcaa ctctcgctca 11440
11441 tcc                                                                                    11442
          |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80
```

(b).

OrfB:

MTTTVYETKSAPQPVGPYSQAVRRGNILSLSGQVALDPATGKVVEGGIAEQTRQVMNNLRNVLESAGASFSDVVTVRIYLTEAADFPGMNSVYAEFLAEPFP
CRTTVFVGLPPGMLIEIDVLAVVE

| 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100

OrfC:

MTLGTARNTVGTRAGLLGPLALAGLAARARASGALDIAIGTPPGQPPRAAVTAAADAMLAGRNQYADPAGVPELRAAVAGRLLASSGVGVDPETEITITAGA
TEGLLVALLTVTDPGDEVVVVEPAFELYRGAVELAGCTRVPAPLDPSGWRLDPERIASVLTPRTAALIVNTPHNPTGRVFTAEELGGLLDLCERRGVTCVVD
EVYADFVFDGGPHVSALGFPRHRPRVVVVGSLSKADEMSGWRLGYCVADPALTTALRHVHERTTFGAAVPLQHGAAVLARAGSGPEQFRNGRDEMVRRLAAM
GFEVTTPEGGWFLLAGTRRLGLRSDALAEGLLVRAGVLVAPATPFFADTGEGQRWIRTTFVKDLATLTKALDAMESFLAEPDGGVWA

| 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100

OrfD:

MTEPKDLTATADTAGTADPNDPTGPAGAAVTRKPDLVTLDEVRAAARRVTEVARRTPLLRCDGLHGGHRVFVKAENLQSTASFKVRGAANALLSLDPRPRSV
VTFSAGNHGAAVALVGRALGVTVTVCMPPGAVAAKVGAVRRYGGEIVFTDDLIGSAAALAAERGCPVLHPFDDPAVIAGQGTTALEIFEEVPHPDVILVPVG
GGGLISGVAAVTAALSDGTRVIGVEPETAPTVGHALRRGTPEPLPVPPRSAADGLNAPRTGALPLAHVRQHAVELVTVGEEDIRRAWPLMAAETRLLVEPSA
AVTVAALRSGRIGLPECATVVLIATGGNTSLGALGDAASSANAVP

| 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100

OrfE:

VTLTVLVVLALGTFAFRLAGVVLNDRLVLTEPVRRAVSVSAIALIFALAVQSAVAEGTAFAGYARLGGVVVGAVLVWRRASFLTVVVAAVATTALLRFLGVA

| 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100

OrfF:

MPGEQRRRRGGGVPPKPGGPGRGEAPGPDGAEVARQIVAIGLGVVAIGVSFGAVAVTSGVPGWMAVALSVFVFAGGAQFMAVSLATAGNPVAAVVAGLLLNA
RHLPFGLALGDMFRPGRGRMLLGAHLLSDESAALALSLPPGPARRRAYWRLAWVLVFAWNGGTVLGVILVDAVGDPAALGLDAANPASLLALTLPALRHREM
RGPVLTGALVCVLTTPFLPAGLPVLLASAGLLLLVPSPFRFRFRPGGRP

| 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100

OrfG:

VKSPDMSRGARLATHGAVSTALALYDDRRLGELVDAGTPAGSGIGGETVLLEVGGTPVFVKRIRLTDLERRPENVRSTANLFGLPLFCHYGVGRIGGPGFGA
WRELAAHTMTTDWVIAGHHEGFPLMHHWRVLPAPDRPLPEELADVERAVAHWGGGPGIRHRIEALQESSASLTLFLEYIPQNLHDWLGARFKAGAEAAERAC
ALVESRLTEGVSFMNTRGLLHFDAHFQNILTDGRRLFFTDYGLALSSRFELARDEADFFDGHRSYDRSYTATHLVNWLTVALHGYGREERAAFIRACAQGVP
PRGVPAAVAAALLRDAPVADVMGDFYRRFQEESRTTPYPLEALRRVGAGGGTTRRRPA

| 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100

OrfH:

MVSQTPVAHTGPPSIRTTSWQWLSVLVLSLSTFVVVTSEMLPVGVLTPMADGLRVTPGTAGFSLTVTGLVTAVTAPAVPRLLGARDRRAVLAAAMVVLALGN
VLTALAHGFGLLVVSRIVLGIGMGAVWGLAAAVAPRLVAPRDIALAVSVTVGGVAAASVVGVPLGTLVSNAFGWRAAFATLAGGAVLLAAGLLLALPRLPRP
ENPAGTGDRTADPSGETRGGDVSLPRGRPVLGGLLLVLLLVTAHFAAYTYVRPVLEERTGLDPGAVALVLLVYGLFGLAGNFAAGSLAVRRARGTVLGLTAG
IALSIALLALFGATAGAAGVAVALWGLTYGGLSVAGQIWMTRAAPHRVEQVTGVYVGVFTGAIALGAFLGGTVVEAAGITTLLWSATALAVVAFGVGLTGGR
PAAVKRHP

| 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100

OrfI:

VNQVHVQEIECLLVLAEELHFGRTAERLGCSQSRVSQLVAGLERRIGVRLVDRTSRSVALSRFGTQFVEEVRPAYEALNTVITQARDRARSGTLGQLRIGFH
GSVYEEVTEAFRRLRAHHRVALVPSEIPLGSPFTAVLEGRVDAVVVELPVNEPTLTTGFRFPPQERLLAVATRHPLAGGDRVHIEELATLDILHPLGDAPEY
WRAARVPPATPGGVPIRSTAGITTVQEGLALVATGDQGLLVCRPLADRAVRDDVRFLPVDGLDEPSQLGLIWREDRTSWQLTTLARLLAEEFARSADAGRAA
GCPSPDAGTDTATGSGTAAGTDTGIGTGIRTGTGTGTGARARTPNAPARV

| 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100

OrfJ:

MSRETSRRPTMRTLISTAFISLDGVVEAPGGEPGYRNSGWTFKEVEFVPEAFEIKGAEQQEATAVLLGRVSYEAFSPVWPDMEEFARYREMPKYVVSSTLTD
ADLVSNWGETTILRSLDEVTALKETEGGPIILHGSASLNRALSDAGLIDRYHLLVHPLLLGAGKRLFSTSDRDAQKLELVEHEVYANGVQKQIFDVVR

| 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100

OrfK:

MTHPFEITLETTLPASPEEVWEAIATGPGIDSWFMGRNEVEAREGGVAAMETGGHREEAVITAYEPGSRFATRTATAADGRFMAFEYLIEGRDRGSTVLRVV
HSGMLGEDWQDEYDALRRGWPFHLHTLREYLTHFAGRTGVPVFAMAPSGGRPVREIRAALTRGLSLAEPVAVGARAHAEPAGLPRFDGEVVWADDERFEVRT
ADGIYTFHHSAAVVLMFHHLFVTGTGTGTDTDGGTDGDGTGSGGATGDGVGDTTVAGDRAEAAWREWLTGLLA

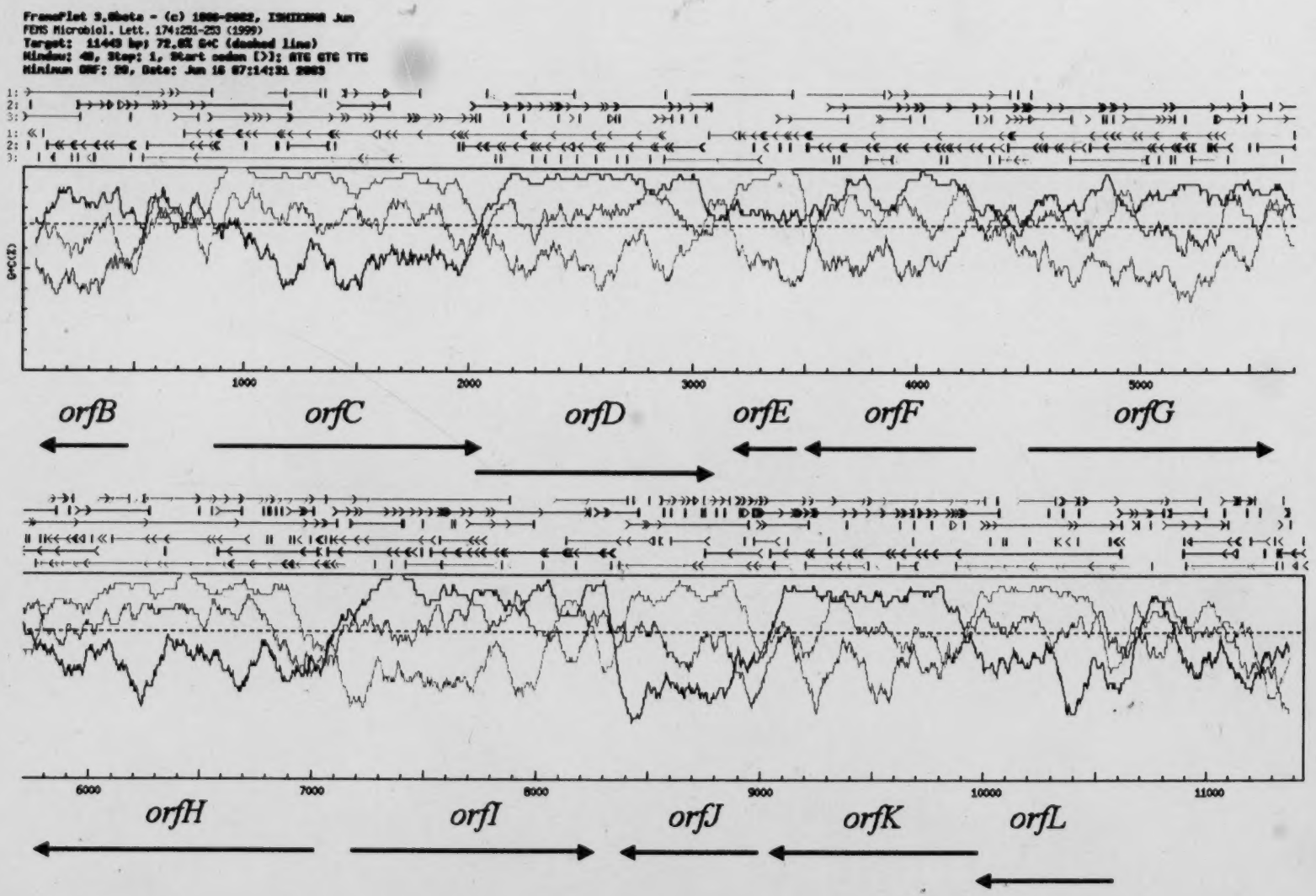
| 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100

OrfL:

VIEEPAAAEASLDPIRSRILAALAEPGSAAMLATRLGLARQKVNYHLKELERHGLVELVEERRKGNVTERVYGATAASYVISPSALAAVSPDPSRSPDQLSA
RWLLALGARLVQEVGSLLTGAARVRRRVATFGIDAEVRFASAADRAAFADELTRTVTALVSRYHDESAPAGRSHRVIVGLHQIPAARPGTEEDAPVSRQQPE
GTRATTAPVASTTSEAPDARQEHRP

| 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100

Figure 18. Prediction of the open reading frames present on the 11.4 kb DNA fragment by FramePlot 3.0 beta. Each colored line represents the potential ORFs starting from each position of the first codon. The program predicts protein-coding regions in *Streptomyces* DNA based on the biased G and C codon preference in the third position of the ORFs. Start codons are indicated by small arrows, while stop codons are indicated by small bars. Both DNA strands were analyzed and shown. Gene designations of putative open reading frames are labeled and the orientation of transcription is indicated by arrows.


FramePlot 3.0beta - (c) 1998-2002, IINUMAYA Jun
FEMS Microbiol. Lett. 174:251-253 (1999)
Target: 11449 bp; 72.6% G+C (dashed line)
Window: 45, Step: 1, Start codon [>]: ATG GTG TTG
Minimum ORF: 20, Date: Jun 16 07:14:31 2003
G+C(%)
1000
2000
3000
4000
5000
orfB
orfC
orfD
orfE
orfF
orfG
6000
7000
8000
9000
10000
11000
orfH
orfI
orfJ
orfK
orfL

frequencies in the codon's third position of the ORFs are in the range from 85.0% to 96.1% (Table 4). Both are in agreement with this characteristically biased codon preference of *Streptomyces*.

Whereas *orfC*, *orfD*, *orfG* and *orfI* are transcribed in the same forward orientation, *orfB*, *orfE*, *orfF*, *orfH*, *orfJ*, *orfK* and *orfL* are oppositely oriented. The details of each newly discovered gene are summarized in Table 4 and also discussed further with their deduced gene products in section III.3.3.

Transposon-flanking sequences were used to determine the orientation and the location of Tn*5* on the 11 kb *Eco*RI fragment of cosmid 6G9. As a result, Tn*5* was found to be present at scattered sites (Figure 19). Of the 30 transposon-containing cosmids which were determined to carry Tn*5* on the 11 kb *Eco*RI fragment of cosmid 6G9, three (**B50**, **A2** and C34) contained Tn*5* on *orfC*, two (**D41** and D35) contained Tn*5* on *orfD*, two (**A23** and D24) contained Tn*5* on *orfE*, three (B26, D6 and **B30**) contained Tn*5* on *orfF*, two (C50 and **B40**) contained Tn*5* on *orfG*, four (A49, A7, **A42** and A14) contained Tn*5* on *orfH*, five (C32, B25, A11, A19 and **B7**) contained Tn*5* on *orfI*, two (**A5** and **D10**) contained Tn*5* on *orfJ*, three (B1, B6 and **B20**) contained Tn*5* on *orfK*, and one (**B42**) contained Tn*5* on *orfL*. Three transposon-containing cosmids (B37, B15 and C9) contained Tn*5* in the intergenic regions. For the above transposon-containing cosmids mentioned, those shown in bold were chosen to construct *S. clavuligerus* mutants with priority given to those carry Tn*5* lies in the same orientation with the identified ORFs and to those carry Tn*5* lies in the 5' to the middle region of the ORFs. However, some

Table 4. Properties of the newly discovered open reading frames.

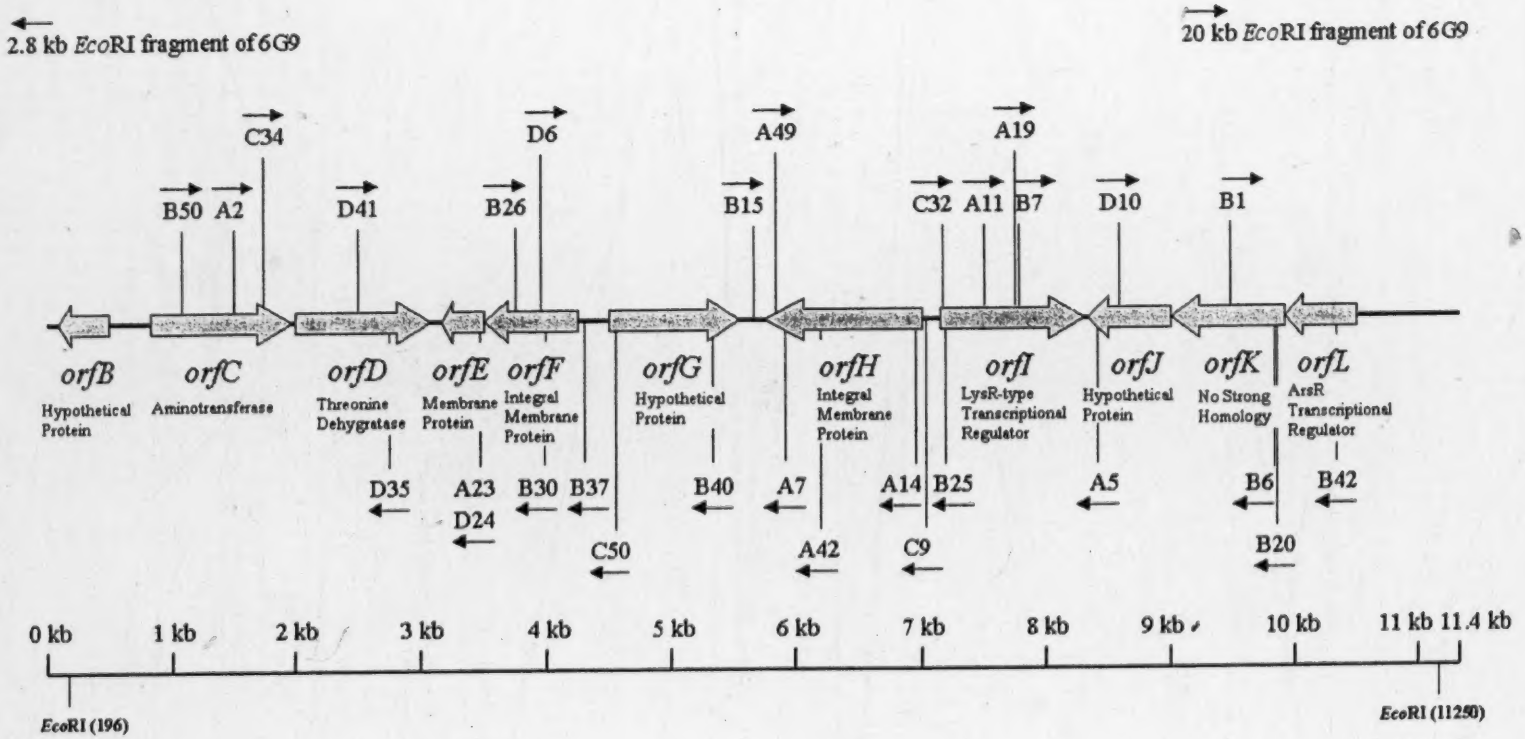
ORFs	ORF Location (nt)[a]		Size (bp)	G + C%[b]	Start Codon	Stop Codon	Transposon-containing Cosmid[c]
	Start	End					
orfB	483	103	381	85.0	ATG	TAA	None
orfC	852	2033	1182	93.7	ATG	TGA	**B50**; **A2**; C34
orfD	2030	3085	1056	92.9	ATG	TGA	**D41**; D35
orfE	3511	3203	309	96.1	GTG	TGA	**A23**; D24
orfF	4269	3508	762	88.6	ATG	TGA	B26; D6; **B30**
orfG	4502	5596	1095	86.8	GTG	TAG	C50; **B40**
orfH	7010	5760	1251	89.4	ATG	TGA	A49; A7; **A42**; A14
orfI	7172	8242	1071	89.6	GTG	TGA	C32; B25; A11; A19; **B7**
orfJ	8969	8367	603	89.6	ATG	TGA	**A5**; **D10**
orfK	9873	9040	834	90.3	ATG	TGA	B1; B6; **B20**
orfL	10559	9870	690	87.4	GTG	TGA	**B42**

[a] The putative open reading frames were identified by FramePlot 3.0 beta by calculating biased usage of G and C nucleotides in the third position of codons and detecting start and stop codons nearby the ORFs.

[b] The G + C percentages of the newly discovered genes were determined using FramePlot 3.0 beta.

[c] The transposon-containing cosmids that contain Tn*5* in each of the genes are listed, and those chosen to construct *S. clavuligerus* mutants are printed in bold.

Figure 19. Arrangement of the newly discovered open reading frames relative to the sites and orientations of transposon insertions. The line in the middle represents the *S. clavuligerus* DNA sequenced, and the grey arrows on the line represent the open reading frames and the direction of their transcription. The orientation of transposon insertions was indicated by small arrows.


2.8 kb EcoRI fragment of 6G9
20 kb EcoRI fragment of 6G9
C34
D6
A49
A19
B50
A2
D41
B26
B15
C32
A11
B7
D10
B1
orfB
Hypothetical Protein
orfC
Aminotransferase
orfD
Threonine Dehygratase
orfE
Membrane Protein
orfF
Integral Membrane Protein
orfG
Hypothetical Protein
orfH
Integral Membrane Protein
orfI
LysR-type Transcriptional Regulator
orfJ
Hypothetical Protein
orfK
No Strong Homology
orfL
ArsR Transcriptional Regulator
D35
A23
D24
B30
B37
C50
B40
A7
A42
A14
B25
C9
A5
B6
B20
B42
0 kb
1 kb
2 kb
3 kb
4 kb
5 kb
6 kb
7 kb
8 kb
9 kb
10 kb
11 kb
11.4 kb
EcoRI (196)
EcoRI (11250)

transposon-containing cosmids chosen may not be the best choices to construct gene disruption mutants due to the fact that it was hard to determine if the Tn*5* inserted within the gene until the start codon for these genes was determined. For *orfB*, since no transposon-containing cosmid that contain Tn*5* in the *orfB* gene was identified, the sequence of the *orfB* gene was obtained by primer walking (see section III.4).

After analyzing the transposon flanking sequences, D24 and A23 seemed to carry Tn*5* in exactly the same position of the 11 kb *Eco*RI fragment of cosmid 6G9, suggesting these two may be duplicates of the same parent cells during recovery process of transformation or may result from two independent transposition events.

III.3.3 Deduced gene products of identified open reading frames

To predict the putative function of the newly discovered genes, their encoded proteins were deduced from the DNA sequence by codon preference analysis and database homology searches. The number of amino acids and the calculated molecular weights of the putative proteins were determined using the DNA Strider 1.2 program. A Kyte-Doolittle hydropathy plot (Kyte and Doolittle 1982) was used to predict the location of putative proteins within the cell. The potential functions of putative proteins were determined by BLASTP searches of the databases at the National Center for Biotechnology Information (NCBI). The BLASTP search ranks each homologous protein against the other homologues for its similarity to the query protein, and the extent of similarity between two sequences can be based on percent sequence identity and a positive matrix score. The putative conserved domains for each ORF were analyzed by an

NCBI Conserved Domain Database (CDD) search. A summary of the results of the database homology searches and the putative properties of the ORFs is shown in Table 5.

III.3.3.1 The *orfB* gene and its deduced product

The *orfB* gene was identified when using primer walking to complete sequence information for the 2.8-kb-*Eco*RI-fragment end of the 11 kb *Eco*RI fragment of cosmid 6G9. The *orfB* gene is 381 bp in length. The G + C frequency in the third position of *orfB* codons is 85.0%. The translational start codon for the *orfB* gene was predicted to be an ATG starting at position 485 and was preceded by a putative ribosome-binding site sequence TCCT.

Deduced from the sequence data obtained, the OrfB protein was predicted to be a 126-amino-acid polypeptide (Figure 17b) with a calculated molecular weight of 13.4 KD. The hydropathy plot of the OrfB protein did not show a concentration of hydrophobic amino acids in any area (data not shown), suggesting that the protein is not membrane associated. The OrfB protein, when compared to those in the current protein sequence databases by BLASTP analysis, showed a high degree of similarity to the derived amino acid sequences of many hypothetical proteins. The strongest homologs were a hypothetical protein PH0854 from *Pyrococcus horikoshii*, a conserved hypothetical protein from *S. coelicolor*, and a hypothetical protein from *Pyrococcus furiosus* (Table 5). NCBI CDD analysis showed that all these hypothetical proteins contained conserved domains from the YER057c/YjgF/UK114 family, and OrfB contained a conserved domain that 98-100% aligned to those present in the proteins belonging to YjgF family

Table 5. Deduced gene products of the ORFs present on the 11 kb *Eco*RI fragment on cosmid 6G9 and their homologous proteins.

ORF[a]	M.W. (KD)[b]	Proteins with highest similarity[c]		NCBI Accession Number	Similarities			
		Organism	Protein & Length[d]		Identities (%)	Positives (%)	Length[e]	*E*-value[f]
OrfB (126)	13.4	*Pyrococcus horikoshii*	Hypothetical protein PH0854 (126)	O58584	49	71	120	7e-28
		S. coelicolor	Conserved hypothetical protein (127)	NP_631776	50	68	126	3e-27
		Pyrococcus furiosus	Hypothetical protein (126)	NP_578397	49	70	119	6e-27
OrfC (393)	40.8	*Arabidopsis thaliana*	Putative aminotransferase (440)	AAL67016	33	52	361	2e-54
		A. thaliana	Putative aminotransferase (404)	AAM67247	34	52	361	2e-53
		Xanthomonas axonopodis	Succinyldiaminopimelate transaminase (382)	NP_642639	34	51	377	5e-52
		P. fluorescens	Hypothetical protein (406)	ZP_00087765	36	51	355	2e-51
		Thermobifida fusca	Hypothetical protein (391)	ZP_00056895	36	49	366	1e-50
		Cytophaga hutchinsonii	Hypothetical protein (383)	ZP_00119500	33	50	366	1e-50
		S. avermitilis	Putative aminotransferase (398)	NP_825694	37	50	383	3e-50
		Pseudomonas syringae	Aminotransferase, classes I and II (382)	NP_791266	33	49	373	3e-50
		Burkholderia fungorum	Hypothetical protein (384)	ZP_00031145	37	52	345	4e-50
		Xylella fastidiosa	Aminotransferase (382)	NP_779669	35	51	371	5e-50
		S. coelicolor	Putative aspartate aminotransferase (396)	NP_627852	37	52	390	6e-50
OrfD (351)	35.5	*Clostridium perfringens*	Threonine dehydratase (402)	NP_562081	33	55	320	4e-43
		Clostridium tetani	Threonine dehydratase (405)	NP_783137	35	54	310	5e-43
		S. coelicolor	Putative threonine dehydratase (325)	NP_625121	43	56	307	5e-43
		Halobacterium sp. NRC-1	Threonine dehydratase (495)	NP_280764	38	52	313	1e-40
		Thermoplasma volcanium	Threonine dehydratase (406)	NP_110709	35	52	308	4e-40

Table 5. Continued...

ORF[a]	M.W. (KD)[b]	Proteins with highest similarity[c]		NCBI Accession Number	Similarities			
		Organism	Protein & Length[d]		Identities (%)	Positives (%)	Length[e]	*E*-value[f]
OrfE (102)	10.5	*S. coelicolor*	Hypothetical protein SC4H8.03 (102)	T35134	37	53	93	1e-10
		Salmonella typhimurium LT2	Putative inner membrane protein (107)	NP_462921	38	58	96	3e-04
		Salmonella enterica subsp. enterica serovar Typhi	Hypothetical protein (81)	NP_458009	39	60	64	0.59
OrfF (253)	25.7	*S. enterica* subsp. enterica serovar Typhi	Putative exported protein (219)	NP_458008	36	56	195	1e-17
		S. coelicolor	Putative membrane protein SC4H8.02 (256)	CAA15869	35	48	217	3e-16
		Thermobifida fusca	Hypothetical protein (243)	ZP_00057788	35	50	191	3e-14
		S. avermitilis MA-4680	Hypothetical protein (302)	NP_823674	35	49	212	4e-14
OrfG (364)	39.9	*S. avermitilis* MA-4680	Hypothetical protein (358)	NP_821993	51	65	356	5e-87
		S. coelicolor A3(2)	Hypothetical protein (357)	NP_624571	47	61	355	8e-77
OrfH (416)	41.5	*Mesorhizobium loti*	Hypothetical protein (475)	NP_106730	36	54	387	3e-43
		M. loti	Hypothetical conserved transmembrane protein (408)	CAD31518	36	54	387	1e-41
		Pseudomonas putida KT2440	Major facilitator family transporter (408)	NP_744791	37	53	396	3e-41
		S. coelicolor A3(2)	Putative integral membrane transport protein (407)	NP_624429	35	54	380	3e-39

Table 5. Continued...

ORF[a]	M.W. (KD)[b]	Proteins with highest similarity[c]		NCBI Accession Number	Similarities			
		Organism	Protein & Length[d]		Identities (%)	Positives (%)	Length[e]	*E*-value[f]
OrfI (356)	38.1	*S. coelicolor* A3(2)	Putative LysR-family transcriptional regulatory protein (297)	CAB65570	40	51	229	3e-30
		Bacillus anthracis str. Ames	Putative *als* operon regulatory protein AlsR (301)	NP_845992	32	53	191	8e-19
		Bacillus cereus ATCC 14579	LysR family transcriptional regulator, (294)	NP_833396	34	53	186	1e-18
		B. anthracis A2012	Bacterial regulatory helix-turn-helix protein, LysR family (193)	NP_657572	32	53	191	5e-18
		Bradyrhizobium japonicum	Transcriptional regulatory protein (301)	NP_772510	29	45	292	6e-18
		S. coelicolor A3(2)	Putative LysR-family transcriptional regulator (344)	CAB66273	33	44	295	1e-17
OrfJ (200)	22.2	*S. coelicolor* A3(2)	Conserved hypothetical protein SC1H10.21 (213)	CAB88165	82	91	192	3e-88
		Amycolatopsis orientalis	Hypothetical protein (205)	CAB45046	68	84	195	6e-75
		S. coelicolor A3(2)	Putative secreted protein (212)	CAC16994	38	55	181	4e-22
		S. avermitilis MA-4680	Hypothetical protein (202)	NP_823169	39	53	177	5e-22
OrfK (277)	29.7	*Bacillus halodurans*	Hypothetical protein BH1534 (138)	F83841	25	42	131	0.051
		Ralstonia solanacearum	Conserved hypothetical protein (134)	NP_521742	26	40	136	0.008

Table 5. Continued...

ORF[a]	M.W. (KD)[b]	Proteins with highest similarity[c]		NCBI Accession Number	Similarities			
		Organism	Protein & Length[d]		Identities (%)	Positives (%)	Length[e]	*E*-value[f]
OrfL (229)	24.4	*B. cereus* ATCC 14579	Transcriptional regulator, ArsR family (207)	NP_832965	36	65	72	8e-08
		B. anthracis str. Ames	Transcriptional regulator, ArsR family (207)	NP_845557	33	63	72	1e-06
		P. horikoshii	Hypothetical protein (185)	NP_143759	39	58	86	5e-06
		P. horikoshii	Hypothetical protein (192)	NP_143761	32	55	83	2e-04

a. The length in amino acids of the deduced gene products is shown in brackets. The number of amino acids of the deduced gene products was determined using Frame Plot 3.0 beta.

b. The molecular weight of the deduced gene products was calculated using the DNA Strider 1.2 program.

c. The proteins are listed in order of increasing *E*-values.

d. Length of the homologous protein in amino acid is given in brackets.

e. The specific length of each protein (in amino acids) showing homology to the *S. clavuligerus* protein used to calculate the percent identical and positive matches during the database searches is given.

f. The expectation value (*E*-value) is a measurement of the probability of finding the same or greater homology in a database search by chance.

(data not shown).

YER057c/YjgF/UK114 is a widely distributed family of proteins with unknown function. Members of this family are identified by sequence similarity, and most of them are hypothetical proteins. The family is named for genes or proteins in yeast (YER057c), *E. coli* (*yjgF*), and goat (UK114). All are ~14 kDa proteins and do not occur as domains of larger proteins.

A variety of biological processes have been reported to be influenced by YjgF proteins. For example, the gene product of *aldR* from *Lactococcus lactis* has been implicated in regulation of isoleucine biosynthesis (Goupil-Feuillerat et al. 1997); the gene product of *yjgF* from *Salmonella typhimurium* may have a role in regulation of the isoleucine and thiamine biosynthetic pathways (Enos-Berlage et al. 1998); the gene product of *yabJ* from *B. subtilis* stimulated adenine-mediated repression of the purine operon (Rappu et al. 1999). However, in cases in which protein products have been studied, no certain biochemical activity has been identified.

Therefore, the *orfB* genes may belong to the *yjgF* family, but its role in *S. clavuligerus* is unknown.

III.3.3.2 The *orfC* gene and its deduced product

The *orfC* gene is 1182 bp in length. The G + C frequency in the third position of *orfC* codons is 93.7%. The translational start for the *orfC* gene was predicted to be an ATG starting at position 852 and was preceded by a putative ribosome-binding site sequence GGAG.

The OrfC protein is a 393-amino-acid polypeptide (Figure 17b) with a calculated molecular weight of 40.8 KD. The hydropathy plot of the OrfC protein did not show a concentration of hydrophobic amino acids in any one area (data not shown), suggesting that the protein is not membrane associated. The polypeptide showed good similarity to a large number of aminotransferases, with the highest homologies to a putative aminotransferases from *Arabidopsis thaliana*, a succinyldiaminopimelate transaminase from *Xanthomonas axonopodis*, and several hypothetical proteins and aminotransferases from other organisms including *S. coelicolor* (Table 5). NCBI CDD analysis showed that the deduced polypeptide of *orfC* and its homologous proteins share a conserved domain characteristic of aspartate/tyrosine/aromatic aminotransferases. In particular, OrfC showed a 94.4% alignment to the consensus sequence in the region of this conserved domain (data not shown).

Since OrfC contains a conserved domain characteristic of aspartate/tyrosine/aromatic aminotransferases, and its homolog from *S. coelicolor* is an aspartate aminotransferse, it is possible that OrfC may function as, or at least catalyzes a similar reaction as, an aspartate aminotransferase. Aspartate aminotransferases are enzymes that catalyze the reversible transfer of an amino group from aspartate to α-ketoglutarate to form glutamate and oxaloacetate, using pyridoxal phosphate as a cofactor. Aminotransferases have been classified into four families (I-IV) based on mostly their amino acid sequence homologies. Aspartate aminotransferases, together with aromatic amino acid aminotransferases, and alanine aminotransferases, belongs to Family I (Mehta et al. 1993). In this family, seven

residues which are directly related to the positioning of the PLP component are conserved. They are Tyr^{70}, Gly^{110}, Asn^{194}, Pro^{195}, Tyr^{225}, Arg^{266} and Gly^{268}. The numbering of these residues is according to the numbering for pig cytosolic aspartate aminotransferase (Jensen and Gu 1996). When compared to the protein sequence of pig cytosolic aspartate aminotransferase, all these seven residues were also found in the deduced OrfC protein (data not shown).

Therefore, the OrfC protein is most likely to function as a PLP-dependent aminotransferase.

The deduced polypeptide of *cvm6* in the newly discovered clavam biosynthetic gene cluster also showed good similarity to a variety of PLP-dependent transaminases (Mosher et al. 1999). In the tentative pathway leading to *5S* clavams proposed by Egan et al., it was suggested that the clavaminic acid is successively decarboxylated and deaminated to produce a branch-point intermediate that is then converted to *5S* clavams (Egan et al. 1997). Therefore, the *cvm6* product may play a role in both the decarboxylation and the deamination of clavaminic acid to form clavam metabolites (Mosher et al. 1999). Now with the identification of another PLP-dependent aminotransferase, it would be interesting to analyze an *orfC* disrupted mutant to assess its involvement in *5S* clavam biosynthesis.

III.3.3.3 The *orfD* gene and its deduced product

The *orfD* gene is 1056 bp in length. The G + C frequency in the third position of *orfD* codons is 92.9%. The translational start codon for *orfD* was predicted to be an ATG

starting at position 2030 and was preceded by a putative ribosome-binding site sequence GGAG. This start codon overlapped the translational stop codon of *orfC*, which suggested that the two open reading frames might be translationally coupled.

The OrfD protein is a 351-amino-acid polypeptide (Figure 17b) with a calculated molecular weight of 35.5 KD. The hydropathy plot of the OrfD protein did not show a concentration of hydrophobic amino acids except for the center of the protein (data not shown), suggesting the protein is cytoplasmic. The polypeptide showed high similarities to a large number of threonine dehydratases including those from *Clostridium perfringens*, *Clostridium tetani*, *S. coelicolor*, *Halobacterium* sp. and *Thermoplasma volcanium* (Table 5). NCBI CDD analysis showed that the deduced polypeptide of *orfD* and its homologous proteins share a conserved domain characteristic of IlvA threonine dehydratase. In particular, OrfD showed an 83.3% alignment to the consensus sequence in the region of this conserved domain (data not shown).

Isoleucine, lysine, homoserine, threonine, and methionine all belong to the aspartate-derived family of amino acids. The enzymes that synthesize amino acids of this family have been well characterized. During production of isoleucine, the first committed enzyme is threonine dehydratase, which is encoded by the *ilvA* gene. Like aspartate aminotransferase, threonine dehydratase also belongs to the PLP-dependent enzymes. It catalyzes the dehydration of threonine (and serine) to yield 2-ketobutyrate (and pyruvate), using pyridoxal phosphate as a cofactor.

The relationship between amino acid catabolism and antibiotic synthesis has been

studied previously. For example, aspartate aminotransferase, threonine dehydratase, and valine dehydrogenase were required for the biosynthesis of tylosin in *Streptomyces fradiae* NRRL 2702. It was found that aspartate aminotransferase was essential for both cell growth and tylosin production (Lee and Lee 1991).

Therefore, it is possible that the OrfC and OrfD proteins function together in the biosynthesis of clavams in *S. clavuligerus* by providing the necessary precursors.

III.3.3.4 The *orfE* gene and its deduced product

The *orfE* gene is 309 bp in length. The G + C frequency in the third position of *orfE* codons is 96.1%. The translational start codon for *orfE* was predicted to be a GTG starting at position 3511 and was preceded by a putative ribosome-binding site sequence GGAG.

The OrfE protein was predicted to be a 102-amino-acid polypeptide (Figure 17b) with a calculated molecular weight of 10.5 KD. The polypeptide showed similarities to a very limited number of proteins in the database. This included a hypothetical protein SC4H8.03 from *S. coelicolor*, a putative inner membrane protein from *Salmonella typhimurium* LT2, and a hypothetical protein from *Salmonella enterica* subsp. enterica serovar Typhi (Table 5). No conserved domain was detected by the NCBI CDD analysis in any of these proteins. However, the hydropathy plot of the OrfE protein showed a stretch of hydrophobic amino acids present in the N-terminus of the protein which could possibly act as a membrane anchor (data not shown), suggesting the OrfE protein was membrane associated.

All of the homologous proteins of OrfE were discovered during genome sequencing, and thus no information regarding the putative functions of OrfE and its homologs could be deduced.

III.3.3.5 The *orfF* gene and its deduced product

The *orfF* gene is 762 bp in length. The G + C frequency in the third position of *orfF* codons is 88.6%. The translational start codon for *orfF* was predicted to be an ATG starting at position 4269 and was preceded by a putative ribosome-binding site sequence GAGG. Like the *orfC* and *orfD* genes, the stop codon of *orfF* overlapped the translational start codon of *orfE*, which suggested that the two open reading frames might be translationally coupled.

The deduced OrfF protein is a 253-amino-acid polypeptide (Figure 17b) with a calculated molecular weight of 25.7 KD. The hydropathy plot of the OrfF protein showed many transmembrane domains (data not shown) most likely involved in inserting the protein into the cytoplasmic membrane. The polypeptide showed good similarities to a putative exported protein from *S. enterica* subsp. enterica serovar Typhi, the putative membrane protein SC4H8.02 from *S. coelicolor*, and hypothetical proteins from *Thermobifida fusca* and *S. avermitilis* MA-4680 (Table 5). NCBI CDD analysis showed that the deduced polypeptide of *orfF* and its homologous proteins share a conserved domain characteristic of predicted AzlC branched-chain amino acid permease for azaleucine resistance. In particular, OrfF showed an 87.4% alignment to the consensus sequence in the region of this conserved domain (data not shown).

The AzlC branched-chain amino acid permease belongs to the branched chain amino acid exporter (LIV-E) Family. The LIV-E family consists of pairs of integral membane proteins. For example, the AzlC and AzlD proteins of *B. subtilis* together comprise an efflux pump for branched chain amino acids, leucine, isoleucine and valine (Belitsky and Gustafsson 1997). Genes encoding AzlC and AzlD were in a single operon. Overexpression of AzlC and AzlD gives rise to resistance to azaleucine. Azaleucine is a toxic leucine analogue, amino acid antibiotic that can be incorporated into proteins by mimicking leucine.

To examine the possibility that the OrfE and OrfF proteins may function as a pair of integral membane proteins like the branched chain amino acid exporter from the LIV-E family, the deduced amino acid sequence of OrfE was compared to the amino acid sequence of AzlD of *B. subtilis* by Pairwise BLAST. However, no significant similarities between OrfE and AzlD were found (data not shown). NCBI CDD analysis showed that the OrfE protein did not contain the conserved domain characteristic of AzlD predicted branched-chain amino acid permeases as other AzlD proteins do (data not shown). Thus the newly discovered OrfE and OrfF proteins of *S. clavuligerus* seems unlikely to function like the AzlC and AzlD proteins of *B. subtilis*.

However, it is interesting to note that the two homologous proteins of OrfE and OrfF from *S. coelicolor*, the hypothetical protein SC4H8.03 and the putative membrane protein SC4H8.02, also lie next to each other on the chromosome of *S. coelicolor*. This suggests that OrfE and OrfF are likely to act together towards a membrane associated common

function.

III.3.3.6 The *orfG* gene and its deduced product

The *orfG* gene is 1095 bp in length. The G + C frequency in the third position of *orfG* codons is 86.8%. The translational start codon for the *orfG* gene was predicted to be a GTG starting at position 4502 and was preceded by a putative ribosome-binding site sequence GGAG.

The deduced OrfG protein is a 364-amino-acid polypeptide (Figure 17b) with a calculated molecular weight of 39.9 KD. The hydropathy plot of the OrfG protein did not show any concentration of hydrophobic amino acids (data not shown), suggesting the protein is not membrane associated. The polypeptide showed a high degree of similarity to a limited number of the proteins in the databases. This included a hypothetical protein from *S. avermitilis* MA-4680 and a hypothetical protein from *S. coelicolor* A3(2). Both were discovered during genome sequencing projects. No conserved domains were found in the OrfG protein and its homologs by NCBI CDD analysis (data not shown), and thus the function of the *orfG* gene in *S. clavuligerus* is unknown.

III.3.3.7 The *orfH* gene and its deduced product

The *orfH* gene is 1251 bp in length. The G + C frequency in the third position of *orfH* codons is 89.4%. The translational start codon for the *orfH* gene was predicted to be an ATG starting at position 7010, but no conventional ribosome-binding site could be found.

The deduced OrfH protein is a 416-amino-acid polypeptide (Figure 17b) with a

calculated molecular weight of 41.5 KD. The hydropathy plot of the OrfH protein showed a number of transmembrane domains (data not shown) capable of inserting the protein into the cytoplasmic membrane. The polypeptide showed similarities to many membrane proteins, with highest degree of homology to a hypothetical protein and a hypothetical conserved transmembrane protein from *Mesorhizobium loti*, a major facilitator family transporter from *Pseudomonas putida* KT2440, and a putative integral membrane transport protein from *S. coelicolor* A3(2) (Table 5). NCBI CDD analysis showed that the deduced polypeptide of the *orfH* gene and its homologous proteins shared a conserved domain characteristic of the AraJ arabinose efflux permease. In particular, OrfH showed a 95.4% alignment to the consensus sequence in the region of this conserved domain (data not shown).

The arabinose regulon of *E. coli* consists of five operons scattered around the chromosome. During a genetic search for arabinose-inducible promoters, the *araJ* operon was identified (Reeder and Schleif 1991). AraJ belongs to a large class of multidrug resistance translocators, and in particular to the major facilitator superfamily (MFS).

MFS is the main class of proteins that utilize protons as symporters or antiporters to transport a variety of molecules across the cytoplasmic membrane of bacteria and eukaryotes. The specificity of these proteins is very broad, and some of their substrates include antibiotics and sugars. Some MFS members, such as TetA from *E. coli*, responsible for resistance to tetracycline, have a narrow substrate specificity. Others, such as the *B. subtilis* membrane protein Bmr, which is structurally similar to tetracycline

transporters (25% sequence identity), do not cause tetracycline efflux. Instead, Bmr causes the efflux of a variety of toxic substances, including such structurally diverse compounds as ethidium bromide, rhodamine and acridine dyes, tetraphenylphosphonium, puromycin, Cm, doxorubicin, and fluoroquinolone antibiotics (Neyfakh et al. 1991). A large number of genes belonging to the MFS have been revealed by complete sequencing of bacterial genomes, but their transport capabilities have not been established yet.

Since the OrfH protein shows homology to multidrug resistance translocators, it likely functions as a transporter that exports antibiotics or sugars from the cell.

III.3.3.8 The *orfI* gene and its deduced product

The *orfI* gene is 1071 bp in length. The G + C frequency in the third position of *orfI* codons is 89.6%. The translational start codon for the *orfI* gene was predicted to be a GTG starting at position 7172, and was preceded by a putative ribosome-binding site sequence AAGG.

The OrfI protein is a 356-amino-acid polypeptide (Figure 17b) with a calculated molecular weight of 38.1 KD. The hydropathy plot of the OrfI protein did not show a concentration of hydrophobic amino acids in any one area of the protein (data not shown), suggesting that the protein is not membrane associated. The polypeptide showed high similarity to many transcriptional regulatory proteins of the LysR family, with the highest degree of homology to a putative LysR-family transcriptional regulatory protein from *S. coelicolor* A3(2), a putative *als* operon regulatory protein AlsR from *Bacillus anthracis* str. Ames, a LysR family transcriptional regulator from *Bacillus cereus* ATCC 14579, a LysR

family bacterial regulatory helix-turn-helix protein from *B. anthracis* A2012, a transcriptional regulatory protein from *Bradyrhizobium japonicum*, and a putative LysR-family transcriptional regulator from *S. coelicolor* A3(2) (Table 5). NCBI CDD analysis showed that the deduced polypeptide of the *orfI* gene and its homologous proteins share obvious conserved domains characteristic of transcriptional regulators of the LysR family. The OrfI protein showed a 99% alignment to the consensus sequence in the region of this conserved domain (data not shown). In particular, the OrfI protein resembles members of the LysR family in the presence of the characteristic HTH motif at the N-terminus, corresponding to the DNA binding regions of transcriptional activators, and a conserved LysR substrate-binding domain was also found in the OrfI protein.

Antibiotic pathway-specific regulatory genes have been found in the biosynthetic clusters in several *Streptomyces* strains. For example, the gene product of *actII-orf4* in the actinorhodin gene cluster (Fernandez-Moreno et al. 1991), and the RedD and RedZ protein in the undecylprodigiosin gene cluster in *S. coelicolor* (White and Bibb 1997). In *S. clavuligerus*, CcaR, a regulatory protein encoded by the *ccaR* gene in the cephamycin biosynthetic gene cluster, was found to control several pathways for secondary metabolites including cephamycins, clavulanic acid and the *5S* clavams. In addition, ClaR, another regulatory protein encoded by the *claR* gene in the clavulanic acid gene cluster, is a LysR-like protein that specifically targets the regulation of clavulanic acid (Paradkar et al. 1998).

Therefore, the *orfI* gene product is most likely to function as a LysR type

transcriptional regulator.

III.3.3.9 The *orfJ* gene and its deduced product

The *orfJ* gene is 603 bp in length. The G + C frequency in the third position of *orfJ* codons is 89.6%. The translational start codon for *orfJ* was predicted to be an ATG starting at position 8367, but no conventional ribosome-binding site could be found.

The OrfJ protein is a 200-amino-acid polypeptide (Figure 17b) with a calculated molecular weight of 22.2 KD. The hydropathy plot of the OrfJ protein did not show a concentration of hydrophobic amino acids in any area of the protein (data not shown), suggesting that the protein is not membrane associated. The OrfJ protein showed significant similarities to many hypothetical proteins, with highest degrees of homology to a conserved hypothetical protein SC1H10.21 from *S. coelicolor*, a hypothetical protein from *Amycolatopsis orientalis*, a putative secreted protein from *S. coelicolor*, and a hypothetical protein from *S. avermitilis* MA-4680 (Table 5). NCBI CDD analysis showed that the deduced OrfJ protein and its homologous proteins share a conserved domain characteristic of FolA dihydrofolate reductase for coenzyme metabolism. In particular, the OrfJ protein showed an 89.8% alignment to the consensus sequence in the region of this conserved domain (data not shown).

Dihydrofolate reductase catalyses the NADPH-dependent reduction of dihydrofolate to tetrahydrofolate, an essential step in synthesis both of glycine and of purines and deoxythymidine phosphate (a precursor of DNA synthesis) (Trimble et al. 1988). Although dihydrofolate reductase is found ubiquitously in prokaryotes and eukaryotes,

and is found in all dividing cells, maintaining levels of fully reduced folate coenzymes, the catabolic steps are still not well understood.

The OrfJ protein does not show similarity to dihydrofolate reductase except for the region of the conserved domain. Therefore, it is unlikely that OrfJ functions as a dihydrofolate reductase, but its possible that OrfJ catalyzes a similar reaction as a dihydrofolate reductase.

III.3.3.10 The *orfK* gene and its deduced product

The *orfK* gene is 834 bp in length. The G + C frequency in the third position of *orfK* codons is 90.3%. The translational start codon for the *orfK* gene was predicted to be an ATG starting at position 9873, and was preceded by a putative ribosome-binding site sequence AGGA.

The OrfK protein is a 277-amino-acid polypeptide (Figure 17b) with a calculated molecular weight of 29.7 KD. The hydropathy plot of the OrfK protein did not show a concentration of hydrophobic amino acids in any area of the protein (data not shown), suggesting that the protein was not membrane associated. OrfK showed low similarities to a limited number of hypothetical proteins, with the highest degrees of homology to a hypothetical protein BH1534 from *Bacillus halodurans* and a conserved hypothetical protein from *Ralstonia solanacearum* (Table 5). NCBI CDD analysis showed that the deduced polypeptide of the *orfK* gene and its homologous proteins share a conserved domain characteristic of unknown function, but this conserved domain is found in a range of bacterial as well as eukaryotic proteins. In particular, the OrfK protein showed a 96.6%

alignment to the consensus sequence in the region of this conserved domain (data not shown).

Therefore, no putative function could be deduced for the *orfK*-encoded protein.

III.3.3.11 The *orfL* gene and its deduced product

The *orfL* gene is 690 bp in length. The G + C frequency in the third position of *orfL* codons is 87.4%. The translational start codon for the *orfL* gene was predicted to be a GTG starting at position 10559, but no conventional ribosome-binding site could be found.

The OrfL protein is a 229-amino-acid polypeptide (Figure 17b) with a calculated molecular weight of 24.4 KD. The hydropathy plot of the OrfJ protein did not show a concentration of hydrophobic amino acids in any area of the protein (data not shown), suggesting that the protein is not membrane associated. OrfL showed similarities to the transcriptional regulators of the ArsR family from *B. cereus* ATCC 14579 and *B. anthracis* str. Ames, and to the hypothetical proteins from *P. horikoshii* (Table 5). NCBI CDD analysis showed that the OrfL protein and its homologous proteins share a conserved domain characteristic of an arsenical resistance operon repressor and similar prokaryotic, metal regulated homodimeric repressors. In particular, OrfL showed a 91% alignment to the consensus sequence in the region of this conserved domain (data not shown).

Arsenic resistance (*ars*) genes are usually arranged in an *ars* operon. These operons share common organization and genes, the two most common forms of which contain

either three (*arsRBC*) or five (*arsRDABC*) genes. The first genes of these operons encode, ArsR, an arsenite-responsive repressor (Wu and Rosen 1991).

Although the *orfL* gene resembles an arsenite-responsive repressor encoding gene in an *ars* operon, the *orfL* flanking genes do not resemble those present in *ars* operons. In addition, within the approximately 800 bp of DNA sequence available upstream of the *orfL* gene, no putative open reading frames could be identified. Therefore, the OrfL protein is unlikely to belong to an *ars* operon, but still may possess similar regulatory functions against an unknown target.

III.3.4 Target sites of the transposition reaction

The target sites of Tn*5* transposition were analyzed in this study. It was reported that Tn*5* insertion causes duplication of a 9-bp target sequence that exhibits base duplication at certain positions (Goryshin et al. 1998). The target sites analysis in this study showed that the Tn*5* insertions characterized were indeed flanked by a duplication of a 9-bp stretch of target DNA. A comparison of 29 transposon insertion sites determined in this study indicated that the transposon exhibited little or no target site specificity. A list of the DR sequences and their immediate neighboring positions is shown in Table 6. To avoid biases that might be introduced, the transposon insert sites of D24 and A23 were listed only once, even though they could have resulted from two independent transposition events.

To compare the Tn*5* target sites analyzed in this study to the characteristics frequently found for Tn*5* target sites, a target consensus sequence for this study was

Table 6. Tn*5* insertion sites.

Transposon-containing Cosmids	Start[a] (nt)	Target Sequences (5'-3')[b]	End (nt)
B50	1094	GGCTCCTGGCG	1102
A2	1499	CGCATGTCTCG	1507
C34	1676	TCCAGCACGGC	1684
D41	2534	GGTGCTGCATC	2542
D35	2725	CCGTCTCGGGC	2717
D24/A23	3485	TCGCCCTGGGG	3477
B26	3675	AGCGCAGGCAG	3683
D6	3965	CGCGTTGAGCA	3973
B30	4036	AGGTCATGGCT	4028
B37	4341	AGGCCGCTTCA	4333
C50	4654	CCAGCAGGACC	4646
B40	5346	CAGCCAGTTCA	5338
B15	5675	GGTGCCGTACG	5684
A49	5839	ACTCCACAACA	5847
A7	5931	TCTATGTGGGT	5923
A42	6253	ACCTATGTGCG	6245
A14	6958	ACCTCTTGGCA	6950
C9	7028	CGTTGATTGGT	7020
C32	7170	GCGGTGAACCA	7178
B25	7211	ACCAGCAGACA	7203
A11	7485	GTCTACGAGGA	7493
A19	7742	GACGCTCGACA	7750
B7	7778	GGAGTACTGGC	7786
A5	8427	TGGTCGAGCAC	8419
D10	8594	TCCTTCAGGGC	8602
B1	9525	GCGTCGTATTC	9533
B6	9816	ACCTCTTCCGG	9824
B20	9827	GCTCGCCGGCA	9835
B42	10393	GACCAGACCGT	10401

[a.] The position of the target sequences in the 11.4 kb DNA fragment sequenced is shown, starting from the first nt of the 9-bp target site.

[b.] Transposon insertion target sequences are underlined, and are shown from 5' to 3' along with 2 bases of flanking DNA sequences.

deduced by adding up the occurrence of each nucleotide at each position of the DR as well as the nucleotides immediately upstream and downstream from the target. The calculated transposon insertion target data were summarized in Table 7. The resulting consensus target site reads G-XCTCNNGGX-A (N = all bases, X = G or C). This consensus target site obviously reflected the high G + C content of *Streptomyces* DNA, and was in accordance with the preferred Tn*5* target sites suggested by the previous studies. These studies showed that the preferred target sites for Tn*5* insertions contain G/C pairs at the ends of the 9-bp duplicated sequences (Lodge et al. 1988).

A consensus target sequence for Tn*5* insertion was suggested by Goryshin et al. after analyzing over 350 Tn*5* insertion target sites of both in *vivo* and in *vitro* transposition, using both chromosomal and plasmid DNA from *E.coli*. The consensus target sequence revealed was A-GNTYWRANC-T (Goryshin et al. 1998), where N = all 4 bases, Y = T or C, W = A or T, and R = A or G. When the two consensus target sequences were compared, preferences appeared to be present at most DR positions except for position 7. It should be noted that the deduced consensus target site in this study may differ from the reported consensus target site because the G + C content of target DNA used for transposition reaction was different. In addition, the consensus sequences are clearly not an absolute requirement, since many of the Tn*5* inserts analyzed by Goryshin et al. have occurred at sequences that were different from the consensus at one or more positions.

III.4 Construction of *S. clavuligerus* mutants

Although sequence information was obtained for the eleven open reading frames, it

Table 7. Calculated transposon insertion target data[a].

	L[b]	1	2	3	4	5	6	7	8	9	R[b]
G	10	11	8	8	4	7	8	13	14	10	7
A	8	3	2	4	3	7	6	5	6	2	10
T	5	1	6	11	6	7	9	6	4	2	4
C	6	14	13	6	16	8	6	5	5	15	8
Total	29	29	29	29	29	29	29	29	29	29	29

[a] The bases at the nine positions of the DR and the two flanking letters of the transposon target sites were calculated.

[b] The left (L) and right (R) bases flanking the target sites.

was not evident from the deduced protein sequences whether any of them might play a role in clavam biosynthesis. To gain further insights into the possible involvement of these genes, mutants were prepared and their clavam biosynthetic abilities were compared to the wildtype *S. clavuligerus*.

To construct *S. clavuligerus* mutants, one or two transposon-containing cosmid(s) from each putative ORF was chosen. Priorities were given to the transposon-containing cosmid(s) in which the disrupted ORFs lay in the same orientation with the *egfp* gene, and in which the transposons were located in the 5' to middle region of the ORFs. The transposon-containing cosmids chosen for construction of *S. clavuligerus* mutants were B50 and A2 (from *orfC*), D41 (from *orfD*), A23 (from *orfE*), B30 (from *orfF*), B40 (from *orfG*), A42 (from *orfH*), B7 (from *orfI*), A5 and D10 (from *orfJ*), B20 (from *orfK*), and B42 (from *orfL*). Subsequently, it became apparemt that transposon-containing cosmid C50 would have been a better choice than B40 for the *orfG* gene, but this only became apparent late in the study when additional sequence analysis made possible the assignment of the start codon for *orfG*. Since no transposon-containing cosmid that contain Tn*5* in the *orfB* gene was identified, no *S. clavuligerus* mutant strain with a disruption in *orfB* was constructed.

Another gene disrupted in this study is *orfA*. The *orfA* gene is located on the 2.8 kb *Eco*RI fragment adjacent to *ceaS1*, and thus it is the first gene beyond the left most end of the paralogue gene cluster. The *orfA* gene was partially sequenced by Jensen et al. (unpublished data), but its role in the biosynthesis of clavams in *S. clavuligerus* has not

been characterized yet. Based on the sequence obtained, the putative protein encoded by *orfA* was predicted to be a glycine/serine hydroxymethyltransferase. A transposon-containing cosmid, IC4, was found to contain Tn*5* in the *orfA* gene. Since the object of this study is to investigate the role of the ORFs flanking the left end of the paralogue gene cluster in clavam biosynthesis, *S. clavuligerus* mutants defective in *orfA* was constructed along with all the other mutants even though *orfA* does not reside on the 11 kb *Eco*RI fragment.

To introduce transposon-containing cosmids into *S. clavuligerus* 3585, the cosmids were first transformed into *E. coli* ET12567 (pUZ8002) by electroporation. The *E. coli* ET12567 (pUZ8002) strain was used as a non-methylating donor for conjugation. It contains a helper plasmid pUZ8002, which is an RK2 derivative with defective *ori*T that supplies the RK2 transfer function. Approximately 100 transformants of each transformation were obtained. *E.coli* ET12567 (pUZ8002) cells that carried the transposon-containing cosmids then served as donors for conjugation of the transposon-insertion cosmids into *S. clavuligerus* 3585 as described above (see section II.4.2). Once introduced into *S. clavuligerus*, the disrupted gene exchanged with the corresponding wildtype region of the chromosome by homologous recombination. Approximately 5-10 potential exconjugants of each conjugation were obtained when the donor *E.coli* ET12567 (pUZ8002) cells contained DNA of B50 or A2 (from *orfC*), D41 (from *orfD*), B7 (from *orfI*), A5 or D10 (from *orfJ*), B20 (from *orfK*), or B42 (from *orfL*). The mutants that arose from three different primary exconjugants of each mutant clone

were patched onto ISP4 agar plates supplemented with Apra. After sporulation, the spores of each exconjugant were harvested, filtered through sterile non-absorbent cotton, serially diluted in sterile water, and plated on ISP-4 master plates to allow for isolated colonies.

The master plates were replicated onto Apra-containing ISP-4 plates and Km-containing ISP-4 plates to ensure that efficient replica plating had been carried out, and determine if the isolated colonies on the master plates had arisen from a double crossover event. In case of a double crossover event, the transposon-disrupted copy of the gene replaced the wildtype gene, rather than a single crossover event that would result in the integration of the transposon-containing cosmid onto the chromosome. If single crossovers were to occur, pWE15 would be able to confer Km resistance to the cells. If double crossovers were to occur, the resulting mutants would only grow on apra-containing ISP-4 plates, but not on Km-containing ISP-4 plates, since the transposon-disrupted gene would have replaced the wildtype gene and the pWE15 would be lost from the cells. All the colonies analyzed were Apra-resistant and Km-sensitive, suggesting that the cosmid derived Km resistance gene was not present. This is most likely due to the large flanking region of the genes to be disrupted, which increases the possibility of the occurrence of a double-crossover event. These colonies were considered *S. clavuligerus* mutants and used for further studies.

S. clavuligerus mutants constructed were named according to the transposon-containing cosmids used for gene disruption. For example, mutants disrupted in *orfC* by transposon-containing cosmid B50 were named as *S. clavuligerus* B50.

Mutants disrupted by the same transposon-containing cosmid but arising from different excojugants were numbered from 1 to 3.

When *E.coli* ET12567 (pUZ8002) cells that contained A23 (from *orfE*), B30 (from *orfF*), B40 (from *orfG*), or A42 (from *orfH*) were used as donors for conjugation, no potential exconjugants could be obtained. After incubating the plates for two weeks, few (<3) potential exconjugants started to grow. However, those colonies were unable to be subcultured onto fresh Apra containing ISP-4 plates, suggesting they were not true mutants. Similar conjugation procedures were repeated three times for *E.coli* ET12567 (pUZ8002) cells that contained A23 (from *orfE*), B30 (from *orfF*), B40 (from *orfG*), or A42 (from *orfH*), however, no potential exconjugants were obtained, suggesting that the genes *orfs E-H* may be essential and that disruption in these genes would be lethal to the bacterial cells.

III.5 Analysis of the *S. clavuligerus* mutants

Disruption of genes would cause deficiencies in their activity. To determine the putative involvement of the disrupted genes in the biosynthesis of clavams in *S. clavuligerus*, the phenotype of the mutants constructed with respect to their production of clavams was compared to *S. clavuligerus* wildtype.

During colony development, no differences in growth or in morphological behavior of the *S. clavuligerus* mutant strains constructed were observed on either Soy solid medium or Soy liquid medium, suggesting that the genes being disrupted (*orfC*, *orfD*, and *orfs I-L*) were not involved in primary metabolism or morphological differentiation.

III.5.1 Qualitative analysis of clavam biosynthesis in *S. clavuligerus* mutants

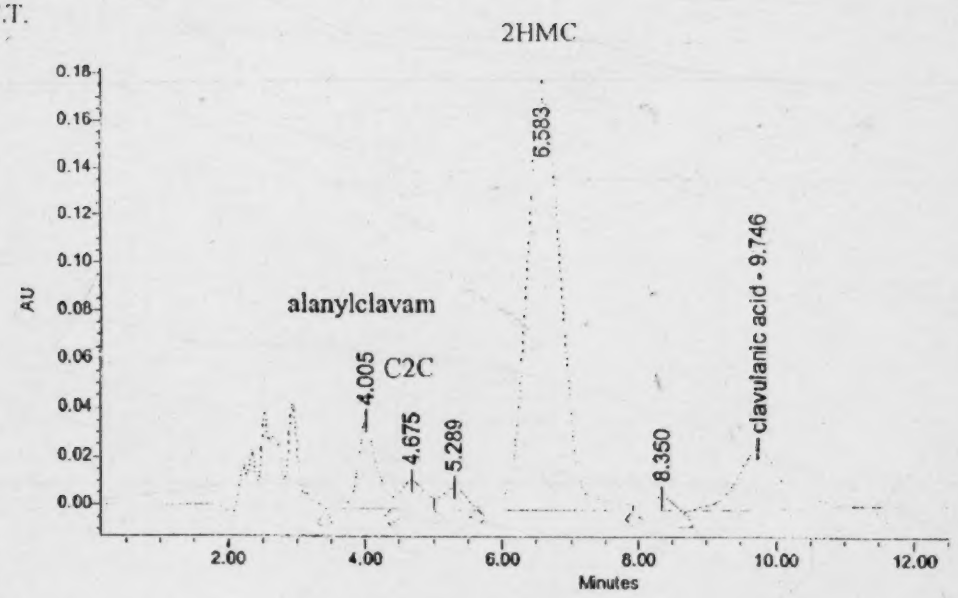
In this study, due to the large number of *S. clavuligerus* mutants to be analyzed, their clavam-producing abilities were first assessed qualitatively by fermentation on solid medium.

Spores of both *S. clavuligerus* wildtype and the mutants were inoculated on Soy solid medium in patches. After sporulation, each patch where organisms grew was transferred into a syringe and frozen at -70°C for 14-18 hr. The syringes were then transferred to room temperature and the liquid was squeezed out of the partially thawed agar. The liquid was collected in an eppendorf tube, and centrifuged to remove any mycelia or particulate matter.

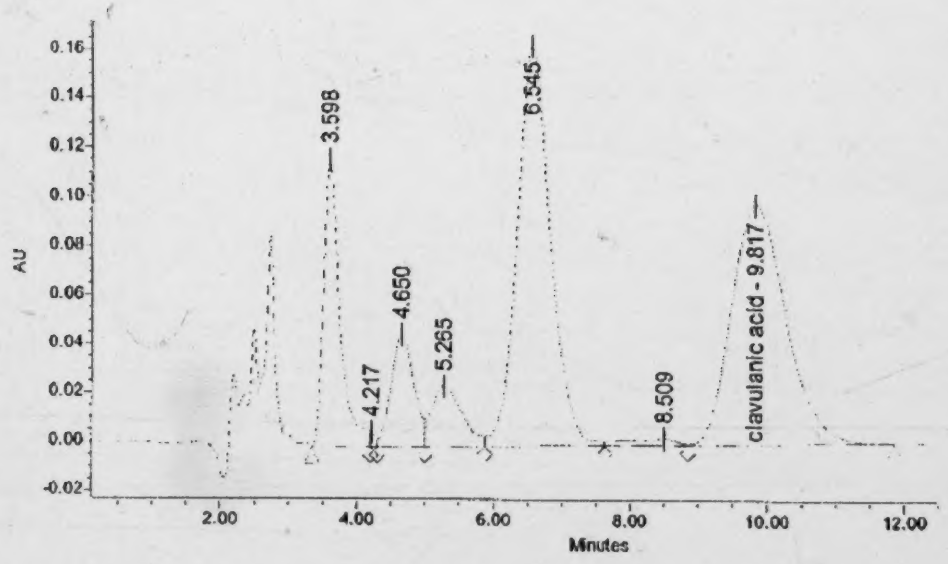
A 100 µl aqueous sample from each agar patch was treated with 25 µl imidazole reagent that reacts specifically with β-lactam-containing compounds, and incubated at room temperature for 15 min to form imidazole-derivatized clavams. A 50 µl sample of each derivatized supernatant was then analyzed by HPLC with detection at 311 nm. Samples without imidazole derivatization were also analyzed by HPLC to identify background peaks, which were subtracted from the peaks of imidazole-derivatized compounds. Peaks representing clavulanic acid and clavam-2-carboxylate in culture broths of *S. clavuligerus* were identified by comparison with authentic standards. 2-Hydroxymethylclavam was identified by its known chromatographic properties relative to those of clavulanic acid and clavam-2-carboxylate. Examples of some of the HPLC chromatograms for these "freeze-squeeze" samples are shown in Figure 20.

Figure 20. Examples of HPLC-chromatograms of "freeze-squeeze" samples by *S. clavuligerus* wildtype and mutant strains on Soy solid medium. One hundred microliters of culture supernatant was mixed with imidazole reagent (25 µl) and incubated at room temperature for 15 min to form imidazole-derived clavams. A 50 µl sample of the derivatized supernatant was then analyzed on a reverse-phase column (Bondapak-C18) with an isocratic buffer system consisting of 0.1 M KH_2PO_4-6% methanol (pH 3.68) and detection at 311 nm. W.T.: wildtype; C2C: clavam-2-carboxylate; 2HMC: 2-hydroxymethylclavam.

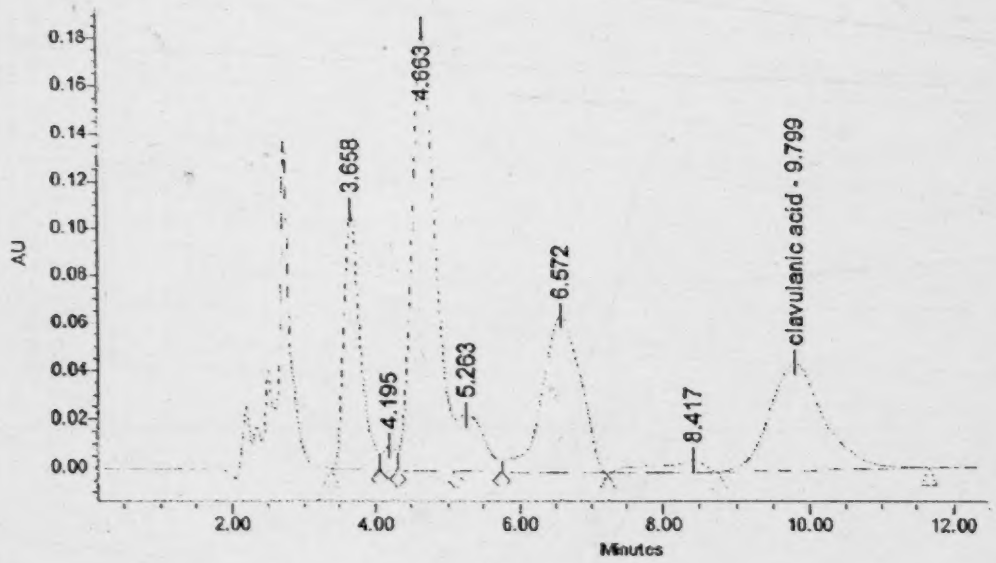
W.T.


2HMC
alanylclavam
C2C
4.005
4.675
5.289
6.583
8.350
clavulanic acid - 9.746
AU
Minutes

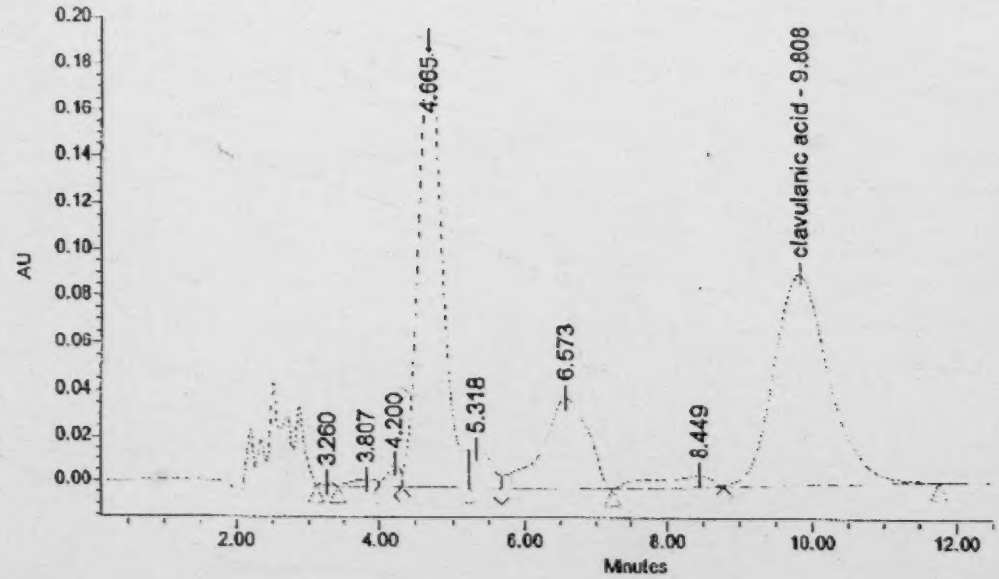
A2-1


3.598
4.217
4.650
5.265
6.545
8.509
clavulanic acid - 9.817
AU
Minutes

D41-1


AU
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
3.658
4.663
4.195
5.263
6.572
8.417
clavulanic acid - 9.799
2.00
4.00
6.00
8.00
10.00
12.00
Minutes

IC4-1


AU
0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
3.260
3.807
4.200
4.665
5.318
6.573
8.449
clavulanic acid - 9.808
2.00
4.00
6.00
8.00
10.00
12.00
Minutes

Although the area of the Soy agar inoculated was similar, the amount of spores used as inoculum was not standardized, and thus the HPLC analysis results obtained here were considered qualitative. A summary of the qualitative analysis results of clavam production of *S. clavuligerus* mutants is shown in Table 8. If all three mutants from a single gene gave a similar fermentation pattern, then only one representative strain was subjected to further quantitative fermentation analysis. When different fermentation patterns were obtained, each corresponding strain was subjected to further quantitative fermentation analysis.

A peak corresponding to clavulanic acid (9.8 min) was seen in all the samples analyzed, indicating that *orfC*, *orfD*, and *orfs I-L* genes were unlikely to be required for clavulanic acid biosynthesis of *S. clavuligerus*.

In terms of changes in *5S* clavam metabolite production, disruption in *orfC* yielded mutants that all failed to produce a peak in the area of alanylclavam (4.0 min), however, still most were able to produce clavam-2-carboxylate (4.7 min) and 2-hydroxymethylclavam (6.6 min) except for *S. clavuligerus* A2-3. Unlike the other *S. clavuligerus* mutants defective in *orfC*, mutant A2-3 did not produce any *5S* clavam metabolites but instead produced an unusually large amount of clavulanic acid (data not shown). Thus, *S. clavuligerus* A2-3 was chosen for further study. Since mutants A2-1, A2-2, B50-1 and B50-2 produced similar patterns of fermentation products (data not shown), only *S. clavuligerus* A2-1 and B50-1 were chosen for further study. In constrast, mutant B50-3 produced a larger proportion of clavam-2-carboxylate and less

Table 8. Qualitative analysis of clavam biosynthesis of *S. clavuligerus* wildtype and mutants[a].

Disrupted Gene	Strain[b]	alanylclavam	C2C[c]	2HMC[c]	CA[c]
------	Wild-type	+	+	+	+
orfC	**A2-1**	- (×)	+	+	+
	A2-2	- (×)	+	+	+
	A2-3	-	-	-	+
	B50-1	- (×)	+	+	+
	B50-2	- (×)	+	+	+
	B50-3	- (×)	+	+	+
orfD	**D41-1**	- (×)	+	+	+
	D41-2	- (×)	+	+	+
	D41-3	- (×)	+	+	+
orfI	**B7-1**	+	+	+	+
	B7-2	+	+	+	+
	B7-3	+	+	+	+
orfJ	**A5-1**	+	+	+	+
	A5-2	+	+	+	+
	A5-3	+	+	+	+
	D10-1	+	+	+	+
	D10-2	+	+	+	+
	D10-3	+	+	+	+
orfK	**B20-1**	+	+	+	+
	B20-2	+	+	+	+
	B20-3	+	+	+	+
orfL	**B42-1**	+	+	+	+
	B42-2	+	+	+	+
	B42-3	+	+	+	+
orfA	**IC4-1**	-	+	+	+
	IC4-2	-	+	+	+
	IC4-3	-	+	+	+

[a.] Presence of peaks was indicated by "+", while absence of peaks was indicated by "-". For those strains in which a new peak at about 3.6 min was present when the peak corresponding to alanylclavam was absent, a (×) marked was added.

[b.] The *S. clavuligerus* mutants chosen for further quantitative studies were printed in bold.

[c.] C2C: clavam-2-carboxylate; 2HMC: 2-hydroxymethylclavam; CA: clavulanic acid.

2-hydroxymethylclavam, and was thus chosen for further quantitative analysis as well.

Disruption in *orfD* gene yielded mutants unable to produce alanylclavam but still able to produce clavam-2-carboxylate, 2-hydroxymethylclavam and clavulanic acid on Soy solid medium. Again, different patterns of fermentation products were observed (data not shown). Much less clavam-2-carboxylate was produced by the mutant D41-2 when compared with mutants D41-1 and D41-3, and therefore *S. clavuligerus* D41-1 and D41-2 were both chosen for further studies.

It should be noted that the disruption of either *orfC* and *orfD* caused complete loss of alanylclavam production on Soy solid medium but a new peak appeared at about 3.6 min in the HPLC chromatograms of almost all the mutants defective in these two genes (Figure 20). The only exception was mutant A2-3, which generated a totally different pattern of fermentation products from others. Although the identity of this new peak was not determined, it could be a metabolite that accumulated when the gene products of *orfC* or/and *orfD* were not present.

Disruption in the *orfI* gene yielded mutants able to produce all of the *5S* clavam and clavulanic acid on Soy solid medium. However, mutant B7-2 and B7-3 produced similar but much lower amounts of 2-hydroxymethylclavam when compared with that generated by the mutant B7-1 (data not shown), and therefore *S. clavuligerus* B7-1 and B7-2 were chosen for further studies.

Disruption in *orfs J-K* generated *S. clavuligerus* mutants that produced patterns of clavam products similar to the wildtype on Soy solid medium. A representative from the

three mutants arising from each Tn*5* insertion was chosen for further studies. They were *S. clavuligerus* A5-1, D10-1, B20-1, and B42-1.

Disruption in *orfA* generated mutants unable to produce the peak in the area of alanylclavam but still able to produce other *5S* clavam metabolites and clavulanic acid. However, unlike mutants disrupted in *orfs C & D*, no new peak was observed. Similar fermentation patterns were present for all of the IC4 mutants constructed. Therefore, only one of them, *S. clavuligerus* IC4-1, was chosen for further quantitative studies.

In total, thirteen *S. clavuligerus* mutants were chosen for further quantitative study, and these are highlighted in bold in Table 8.

III.5.2 Quantitative analysis of clavam biosynthesis in *S. clavuligerus* mutants

To examine the biosynthesis of clavams by *S. clavuligerus* mutants constructed in this study in a more quantitative way, seed cultures of both *S. clavuligerus* wildtype mutants were inoculated from frozen glycerol stocks of spores. The cultures were incubated on a 28°C shaker at 250 rpm for approximately 40 hr in TSBS medium. Seed cultures were used to inoculate 25 mL amounts of Soy medium to 2% (v/v), dispensed in 125 mL flasks. The Soy cultures were then incubated on a 28°C shaker at 250 rpm for up to 96 hr. Similar fermentations on Soy medium were conducted twice for all of the *S. clavuligerus* mutants and the wildtype strains. For each fermentation, the cultures were harvested at the 72-hr and 96-hr time points, and the culture supernatants were used for *β*-lactam bioassays and HPLC analysis.

III.5.2.1 *β*-lactam antibiotic bioassay

Bioassays were conducted by the agar plate diffusion method. Twenty-five microliters of culture supernatant of each strain was applied to a sterile 6 mm paper filter disk set on the surface of agar inoculated with the appropriate indicator organism.

The amount of cephamycin C produced was assayed by using *E. coli* ESS as the indicator organism seeded into TSB agar. The amount of clavulanic acid produced was determined by examining β-lactamase inhibitory activity and assayed by using *K. pneumoniae* as the indicator organism seeded into TSB agar containing 6 μg/mL penicillin G. Duplicate samples were applied to TSB agar lacking penicillin G to account for background levels of inhibition caused by antibiotic contained within the supernatants. The amount of alanylclavam produced was assayed by using *Bacillus* sp. ATCC 27860 as the indicator organism seeded into Davis minimal media. Duplicate samples were also applied to Davis minimal media containing 200 μg/mL L-methionine to counter act the antimetabolite activity of alanylclavam and thereby allow the background zone of inhibition due to antibiotic contained within the supernatants.

A summary of the bioassay results for the first fermentation on Soy medium is shown in Table 9. Since culture supernatants were assayed from 72 hr and 96 hr cultures of all mutants and the wildtype strains grown on Soy medium, only the larger of the two zone sizes was considered. This is because different samples may reach their highest amounts of metabolite production at different time-point, and thus the larger value should better reflect their productivity.

Supernatant from the *S. clauligerus* wildtype and mutants caused very similar sized

Table 9. *β*-lactam bioassay results for *S. clavuligerus* wildtype and mutants.

Disrupted Gene	Strain	Zones of Inhibition (mm)[ab]					
		Cephamycin		Clavulanic Acid[c]		Alanylclavam[c]	
		72 hr	96 hr	72 hr	96 hr	72 hr	96 hr
------	W.T.	29	33	26 (6)	30 (11)	21 (10)	33 (13)
orfC	A2-1	29	32	25 (11)	29 (9)	9 (6)	8 (8)
	A2-3	27	32	28 (11)	32 (10)	11 (7)	11 (10)
	B50-1	30	34	28 (12)	31 (12)	13 (10)	12 (10)
	B50-3	30	33	24 (9)	30 (9)	12 (11)	13 (8)
orfD	D41-1	27	32	13 (7)	23 (10)	12 (9)	9 (7)
	D41-2	30	33	14 (6)	11 (6)	11 (10)	6 (6)
orfI	B7-1	26	30	30 (12)	26 (10)	14 (9)	15 (12)
	B7-2	29	32	29 (12)	31 (13)	21 (11)	17 (11)
orfJ	A5-1	30	34	21 (9)	28 (9)	17 (7)	32 (10)
	D10-1	26	29	23 (9)	27 (10)	23 (12)	31 (13)
orfK	B20-1	30	34	29 (12)	29 (10)	24 (11)	33 (12)
orfL	B42-1	30	33	29 (13)	31 (12)	20 (7)	26 (9)
orfA	IC4-1	29	34	28 (12)	32 (14)	13 (11)	11 (10)

[a] Twenty five microliter amounts of culture supernatants were applied to filter paper disks. Zones of inhibition were measured after incubation for about 20 hr. All the data shown are in millimeters.

[b] Both the 72 hr and 96 hr samples were assayed, and the results were listed under columns "72" and "96", respectively.

[c] The numbers inside the brackets represent zones of inhibition observed on the control plates: in the absence of penicillin G for clavulanic acid or in the presence of L-methionine for alanylclavam.

zones of inhibition on cephamycin C bioassay plates, indicating that the genes (*orfs C, D, I-L & A*) were unlikely to be involved in the biosynthesis of cephamycin.

For clavulanic acid bioassay results, all the mutants except *S. clavuligerus* D41 produce similar amounts of clavulanic acid to the wildtype. In particular, *S. clavuligerus* D41-1 and D41-2 mutants produce approximately 70% and half of the wildtype level clavulanic acid, respectively, indicating that the mutants with defects in *orfD* had a reduced ability to produce clavulanic acid.

For alanylclavam bioassay results, the mutants with defects in genes *orfC*, *orfD* and *orfA* produced much less alanylclavam when compared to the wildtype *S. clavuligerus*, suggesting these mutants had very little or no ability to synthesize alanylclavam. The mutants defective in the *orfI* gene, Mutants B7-1 and B7-2, produce approximately half of wildtype level alanylclavam, suggesting that the mutants defective in *orfI* had reduced ability to produce alanylclavam. In contrast, mutants with defects in *orfs J-L* produced wildtype level of alanylclavam, suggesting that these genes were unlikely to be involved in the biosynthesis of alanylclavam.

The bioassay for the second fermentation on Soy medium showed similar results for cephamycin C and clavulanic acid production. *S. clavuligerus* mutants produced wildtype levels of cephamycin C and clavulanic acid expect for D41 mutants, which produced approximately only half of the wildtype level of clavulanic acid (data not shown). However, production of alanylclavam for the second fermentation on Soy medium was very poor even by the *S. clavuligerus* wildtype, although similar cultivation procedures

were carried out (data not shown). It has been shown in the previous studies that *5S* clavam production is quite inconsistant and slight differences in agitation or aeration can result in dramatic changes in the amounts of *5S* clavam produced (Thai et al. 2001). Thus, the alanylclavam production of the *S. clavuligerus* mutants was not reproducible, and only dramatic and consistent changes in *5S* clavam production were considered to be of significance.

III.5.2.2 HPLC analysis

To further characterize clavam production by *S. clavuligerus* mutants constructed in this study, culture supernatants were treated with imidazole and analyzed by HPLC.

For the culture supernatants of the first fermentation on Soy medium, samples taken at both 72-hr and 96-hr time points were analyzed by HPLC as described in section III.5.1. HPLC analysis data for the first fermentation of the *S. clavuligerus* mutants grown on Soy medium are summarized in Table 10. Data are expressed as approximate percentages based on peak areas of the amount of metabolite seen in the wildtype strain grown on Soy medium. Examples of some of the HPLC chromatograms for the 96 hr samples are shown in Figure 21.

The analyses of clavam production of the second fermentation by the *S. clavuligerus* wildtype and mutants were conducted on the same instrument but equipped with an Xterra reverse phase column. Only the 96 hr samples were analyzed. Imidazole-derivatized culture supernatants (5 µl) were analyzed at a flow rate of 0.25 ml/min. The mobile phase consisted of solvent A (10 mM ammonium bicarbonate, pH 10)

Table 10. HPLC analysis of clavulanic acid and other clavam metabolites production by *S. clavuligerus* wildtype and mutants on Soy medium (I).

Disrupted Gene	Strain	Amount of metabolites produced (% of wild-type)					
		C2C		2HMC		CA	
		72	96	72	96	72	96
------	W.T.	100	100	100	100	100	100
orfC	A2-1	12	6	92	66	510	197
	A2-3	0	5	91	70	72	88
	B50-1	251	236	171	244	665	419
	B50-3	50	81	74	78	294	202
orfD	D41-1	0	89	85	82	57	82
	D41-2	0	11	78	55	41	73
orfI	B7-1	117	154	104	127	72	90
	B7-2	87	15	106	65	915	398
orfJ	A5-1	139	174	99	151	317	268
	D10-1	126	194	97	112	568	550
orfK	B20-1	134	229	83	121	881	744
orfL	B42-1	110	148	92	120	174	195
orfA	IC4-1	178	242	106	131	702	612

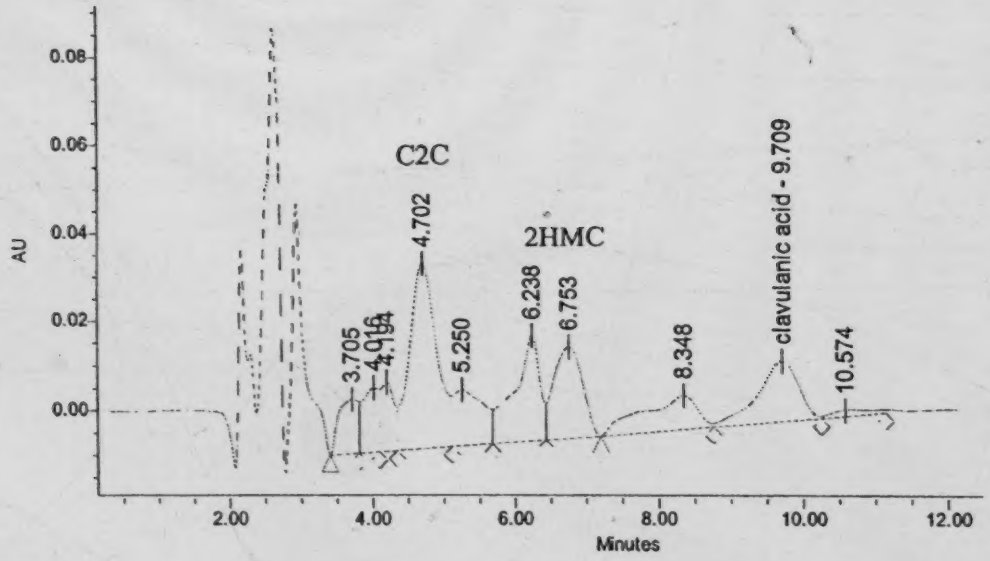
[a] Metabolite production was estimated by HPLC at both 72 and 96 hr. C2C: clavam-2-carboxylate; 2HMC: 2-Hydroxymethylclavam.

[b] For each metabolite, the amount of production seen in the wildtype when grown on Soy medium was assigned a value of 100%.

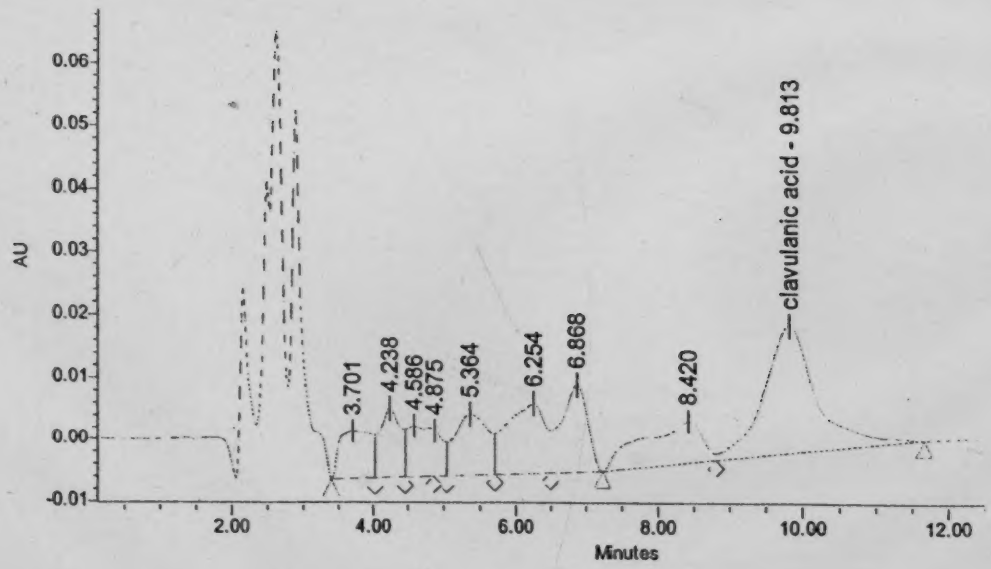
[c] The peak area numbers of the clavam metabolites produced by the wildtype *S. clavuligerus*: C2C (72 hr) 379071; C2C (96 hr) 1096762; 2HMC (72 hr) 366364; 2HMC (96 hr) 569832; clavulanic acid (72 hr) 604277; clavulanic acid (96 hr) 531729.

Figure 21. Examples of HPLC chromatograms of clavulanic acid and clavam metabolites production by *S. clavuligerus* wildtype and mutant strains on Soy medium (I). All are chromatograms of 96 hr samples. One hundred microliters of culture supernantant was mixed with imidazole reagent (25 µl) and incubated at room temperature for 15 min to form imidazole-derived clavams. A 50 µl sample of the derivatized supernatant was then analyzed on a reverse-phase column (Bondapak-C18) with an isocratic buffer system consisting of 0.1 M KH_2PO_4-6% methanol (pH 3.68) and detection at 311 nm. W.T.: wildtype; C2C: clavam-2-carboxylate; 2HMC: 2-hydroxymethylclavam.

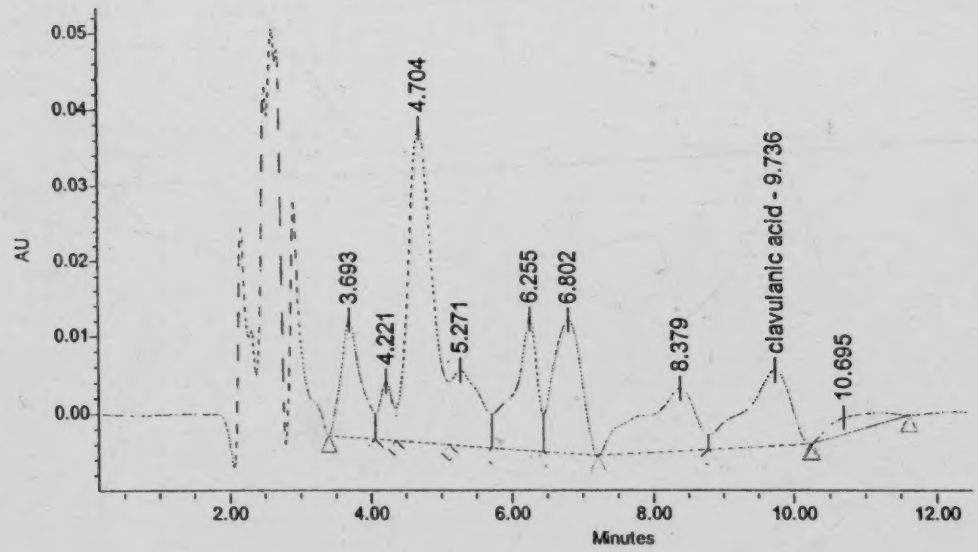
W.T.


AU
0.08
0.06
0.04
0.02
0.00
C2C
2HMC
4.702
3.705
4.016
4.194
5.250
6.238
6.753
8.348
clavulanic acid - 9.709
10.574
2.00
4.00
6.00
8.00
10.00
12.00
Minutes

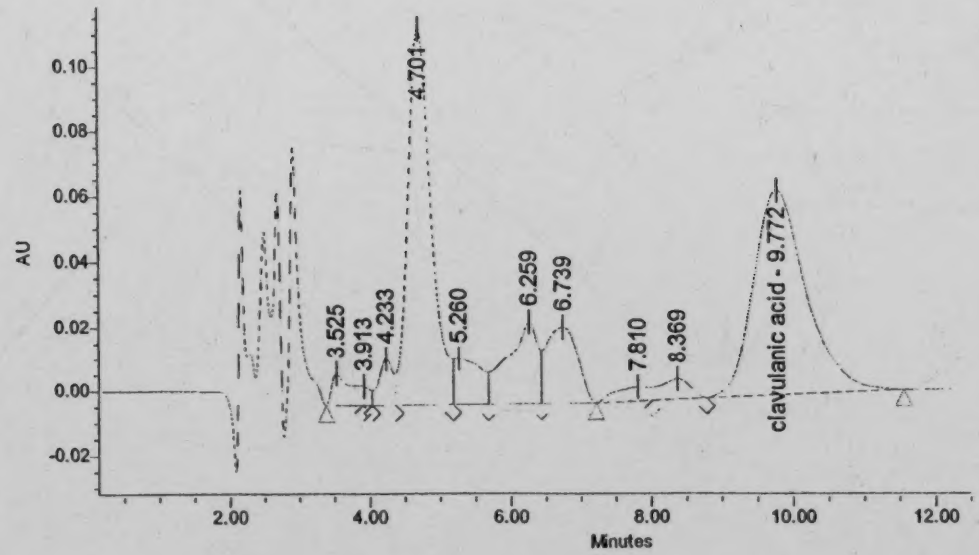
A2-1


AU
0.06
0.05
0.04
0.03
0.02
0.01
0.00
-0.01
3.701
4.238
4.586
4.875
5.364
6.254
6.868
8.420
clavulanic acid - 9.813
2.00
4.00
6.00
8.00
10.00
12.00
Minutes

D41-1


AU
0.05
0.04
0.03
0.02
0.01
0.00
3.693
4.221
4.704
5.271
6.255
6.802
8.379
clavulanic acid - 9.736
10.695
2.00
4.00
6.00
8.00
10.00
12.00
Minutes

IC4-1


AU
0.10
0.08
0.06
0.04
0.02
0.00
-0.02
3.525
3.913
4.233
4.701
5.260
6.259
6.739
7.810
8.369
clavulanic acid - 9.772
2.00
4.00
6.00
8.00
10.00
12.00
Minutes

and solvent B (acetonitrile), used in a gradient system as follows: 100% solvent A for 5 min, linear gradient to 85% solvent A over 10 min, 85% solvent A for 5 min, linear gradient to 100% solvent A over 1 min, 100% solvent A for 9 min. Eluant was monitored at 311 nm to detect the imidazole derivatives. Peaks representing clavulanic acid and clavam-2-carboxylate in culture broths of *S. clavuligerus* were identified by comparison with authentic standards. 2-Hydroxymethylclavam was identified by its known chromatographic properties relative to those of clavulanic acid and clavam-2-carboxylate. HPLC analysis data for the second fermentation of the *S. clavuligerus* mutants grown on Soy medium are summarized in Table 11. Again, data are expressed as approximate percentages of the amount of metabolite seen in the *S. clavuligerus* wildtype grown on Soy medium. Examples of some of the HPLC chromatograms are shown in Figure 22.

The two independent sets of fermentations on Soy medium showed that, in general, lesser amounts of all the clavams were produced in the second set of fermentations, and mutants that are defective in *orfD* tended to produce lesser amounts of all of the clavams. Mutants with defects in *orfs I-L*, *orfC*, *orfD* or *orfA* produced various amounts of clavam-2-carboxylate, 2-hydroxymethylclavam and clavulanic acid. Thus no constant and significant changes in clavam production can be concluded from the mutations based on these preliminary analyses.

III.6 Confirmation of the *S. clavuligerus* mutants by Southern analysis

To confirm that the Tn*5* disrupted copies of the genes have replaced the wildtype genes, genomic DNA from the *S. clavuligerus* wildtype and mutant strains was prepared

Table 11. HPLC analysis of clavulanic acid and other clavam metabolites production by *S. clavuligerus* wildtype and mutants on Soy medium (II).

Disrupted Gene	Strain	Amount of metabolites produced (% of wild-type)		
		C2C	2HMC	CA
------	W.T.	100	100	100
orfC	A2-1	166	74	56
	A2-3	463	189	407
	B50-1	280	265	173
	B50-3	249	233	59
orfD	D41-1	0	0	59
	D41-2	0	0	4
orfI	B7-1	197	82	85
	B7-2	332	122	216
orfJ	A5-1	166	15	49
	D10-1	151	62	70
orfK	B20-1	0	0	94
orfL	B42-1	337	441	249
orfA	IC4-1	0	0	138

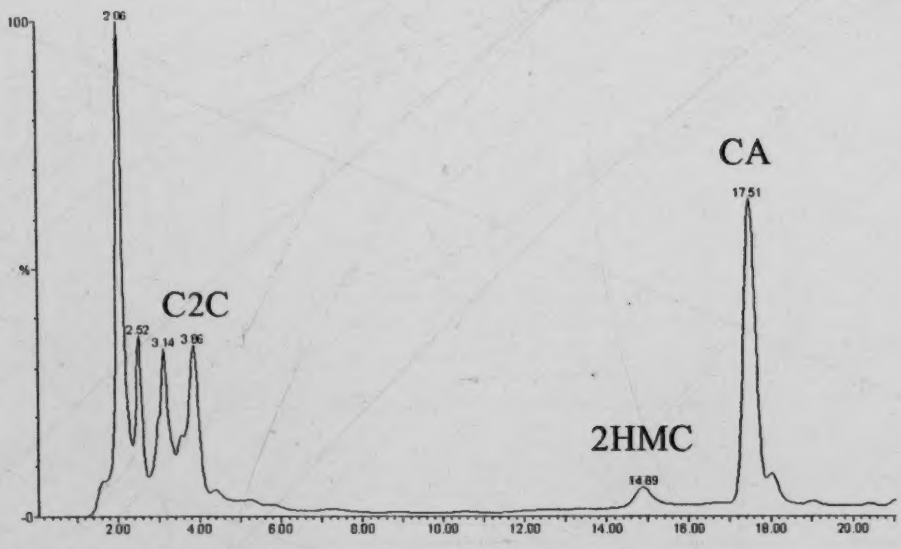
[a] Metabolite production was estimated by HPLC at only 96 hr. C2C: clavam-2-carboxylate; 2HMC: 2-Hydroxymethylclavam.

[b] For each metabolite, the amount of production seen in the wildtype when grown on Soy medium was assigned a value of 100%.

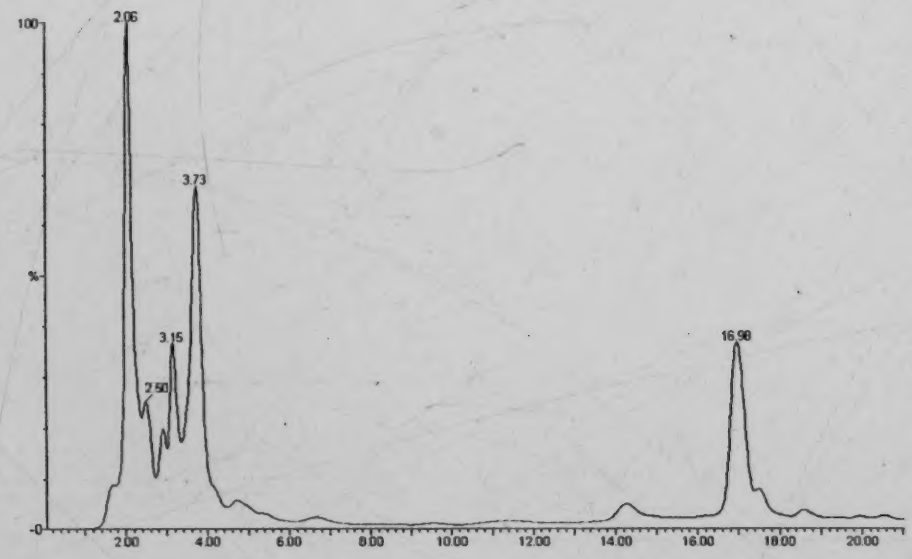
[c] The peak area numbers of the clavam metabolites produced by the wildtype *S. clavuligerus*: C2C 14755; 2HMC 3661; clavulanic acid 44103.

Figure 22. Examples of HPLC chromatograms of clavulanic acid and clavam metabolites production by *S. clavuligerus* wildtype and mutant strains on Soy medium (II). The analyses were conducted on the same instrument but equipped with an Xterra reverse phase column. Imidazole-derivatized culture supernatants (5 µl) were analyzed at a flow rate of 0.25 ml/min. The mobile phase consisted of solvent A (10 mM ammonium bicarbonate, pH 10) and solvent B (acetonitrile), used in a gradient system as follows: 100% solvent A for 5 min, linear gradient to 85% solvent A over 10 min, 85% solvent A for 5 min, linear gradient to 100% solvent A over 1 min, 100% solvent A for 9 min. Eluant was monitored at 311 nm to detect the imidazole derivatives. W.T.: wildtype; C2C: clavam-2-carboxylate; 2HMC: 2-hydroxymethylclavam; CA: clavulanic acid.

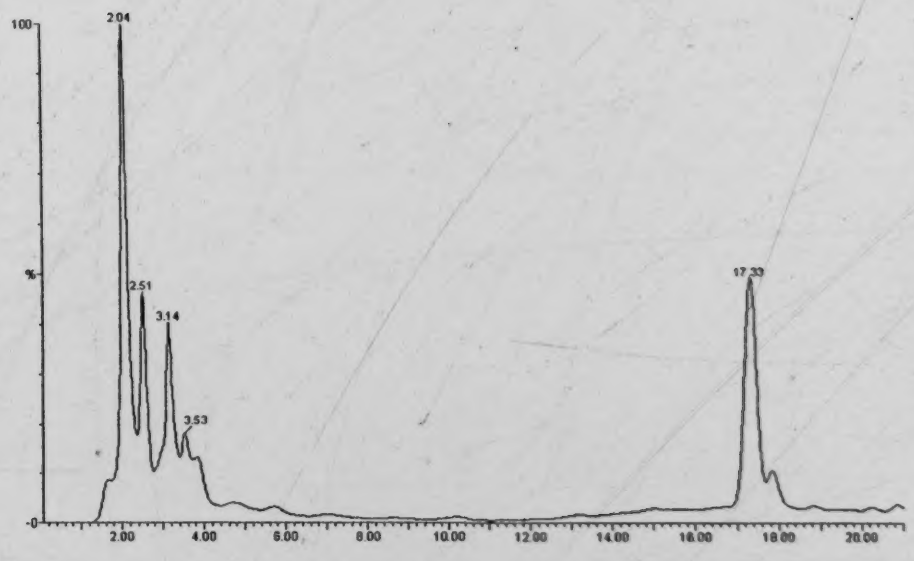
W.T.


100
%
-0
2.06
2.52
3.14
3.86
14.89
17.51
C2C
2HMC
CA
2.00
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00

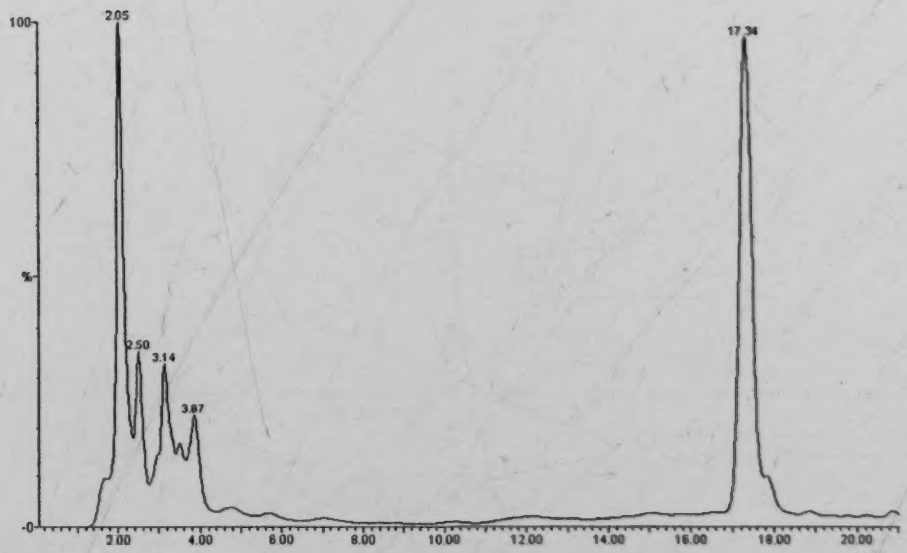
A2-1


100
%
-0
2.06
2.50
3.15
3.73
16.98
2.00
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00

D41-1


100
%
0
2.04
2.51
3.14
3.53
17.33
2.00
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00

IC4-1


100
%
0
2.05
2.50
3.14
3.87
17.34
2.00
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00

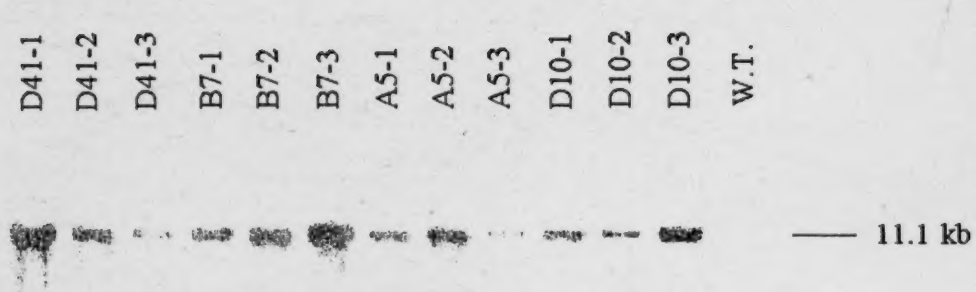
and examined by Southern analysis.

Genomic DNA from all of the 27 *S. clavuligerus* mutants constructed in this study, as well as *S. clavuligerus* wildtype, was prepared, and then digested with *Nco*I. Digested genomic DNA was separated by agarose gel electrophoresis, followed by transfer to a nylon membrane. The membranes were then probed with a ^{32}P-labeled 3.4 kb Tn*5* fragment, which was released from pQM5062 by *Pvu*II digestion. If the wildtype genes were replaced by the Tn*5* disrupted genes, the probe should hybridize to the Tn*5* containing *Nco*I fragments of the mutants, while no hybridization would be seen for the *S. clavuligerus* wildtype. The Southern hybridization results are shown in Figure 23.

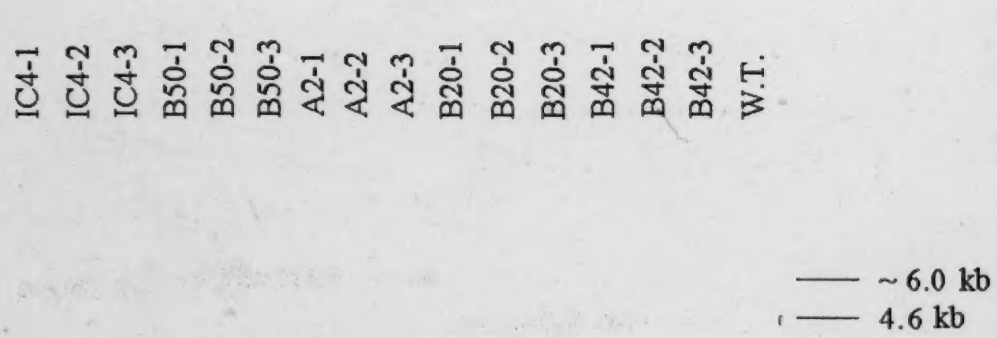
Based on the Tn*5* sequence, it was known that no *Nco*I restriction site is present on Tn*5*, and thus *Nco*I digestion would not separate a Tn*5* hybridized DNA fragment. Four *Nco*I restriction sites were found on the 11.4 kb DNA fragment sequenced. They were at positions 1765, 9432, 9723 and 10906 (Figure 18). Mutants D41, B7, A5, and D10 all contained the 3.4 kb Tn*5* fragment on the 7.7 kb *Nco*I fragment, and thus resulted in a 11.1 kb fragment hybridized to the probe (Figure 23A). Mutants B20 and B42 contained Tn*5* on the 1.2 kb *Nco*I fragment, and thus resulted in a 4.6 kb fragment hybridized to the probe (Figure 23B). Mutants IC4, B50 and A2 contained Tn*5* on the 1765 bp *Eco*RI-*Nco*I fragment, and the other *Nco*I site is somewhere on a unsequenced region of the 2.8 kb *Eco*RI fragment on cosmid 6G9. It was shown that the probe hybridized to an approximately 6 kb fragment in Mutants IC4, B50 and A2 (Figure 23B). No hybridization was seen in the lanes of *S. clavuligerus* wildtype. These results confirmed that the

Figure 23. Southern hybridization analysis of genomic DNA from wild-type *S. clavuligerus* and 27 *S. clavuligerus* mutants disrupted with Tn*5*. Genomic DNA was digested with *Nco*I and separated on two 0.8% agarose gels. After transfer to a nylon membrane, the blots were probed with a 3.4-kb Tn*5* fragment. W.T.: wildtype.

A.


D41-1
D41-2
D41-3
B7-1
B7-2
B7-3
A5-1
A5-2
A5-3
D10-1
D10-2
D10-3
W.T.
11.1 kb

B.


IC4-1
IC4-2
IC4-3
B50-1
B50-2
B50-3
A2-1
A2-2
A2-3
B20-1
B20-2
B20-3
B42-1
B42-2
B42-3
W.T.
~ 6.0 kb
4.6 kb

wildtype genes had been replaced by the Tn5 disrupted copy of genes.

IV. DISCUSSION

While the details of the clavulanic acid biosynthetic pathway have begun to emerge, little is known about the production of the *5S* clavam metabolites by *S. clavuligerus*. Previous studies showed that three gene clusters exist for the biosynthesis of clavam metabolites in *S. clavuligerus*, one directed towards clavulanic acid production and the remaining two towards the other four *5S* clavams. Clavulanic acid and *5S* clavams share the early biosynthetic steps up to and including the formation of clavaminic acid. Paralogue genes have been found for all of the early biosynthetic genes. This includes the paralogue for *cas2*, which has been known since the discovery of the *cas* genes (Salowe et al. 1991) and which is located in the clavam gene cluster, and the paralogues for *ceaS2*, *bls2*, *pah2* and *oat2*, which form the paralogue gene cluster (Jensen et al. 2003b; Tahlan et al. 2003).

The characterization of six *cvm* genes around *cas1*, the paralogue of *cas2*, opened new windows for the study of *5S* clavam biosynthesis in *S. clavuligerus*. Mutants defective in *cvm1* and *cvm4/5* caused a generalized loss of *5S* clavam production (Mosher et al. 1999). However, disruption of structural genes involved in the biosynthesis of particular *5S* clavams should result in selective loss of those metabolites. In this way, details of the biosynthesis of various *5S* clavam metabolites will be deduced, just as has been done for the clavulanic acid biosynthetic pathway. However, so far, no such structural genes for particular *5S* clavams have been found. Therefore, enzymes catalyzing the biosynthesis of particular *5S* clavam metabolites are still unrevealed.

Since clustering of metabolite biosynthetic genes is an invariable rule in *Streptomyces* spp., the most logical way to search for the biosynthetic genes associated with the formation of particular *5S* clavams is to look in the flanking regions of the related genes. To do this, two separate directions were considered. They were, further analysis of the flanking regions of *cas1*, and of the flanking regions of the newly discovered paralogue gene cluster. The objective of this study was to analyze the *ceaS1* side of the genetic flanking regions of the paralogue gene cluster to identify putative biosynthetic genes for *5S* clavam metabolites in *S. clavuligerus*, and determine their potential involvement in clavam biosynthesis as well.

The conventional gene disruption approach used within the Jensen group is laborious and time consuming. Illuminated by the recent reports that an *in vitro* transposition system based on Tn*5* functioned well with *S. coelicolor* DNA and gave random insertions across large stretches of DNA (Gehring et al. 2000), the group planned to explore *in vitro* transposon mutagenesis on *S. clavuligerus*. To date, transposon mutagenesis has been applied only on a few *Streptomyces* spp., including *S. coelicolor* (Gehring et al. 2000), *S. avermitilis* (Ikeda et al. 1993; Weaden and Dyson 1998), *S. tendae* (Engel et al. 2001), *S. roseosporus* (Baltz et al. 1997), and *S. lividans* (Herron et al. 1999), and most of these applications were based on *in vivo* transposition systems. Therefore, this study represents the first application of *in vitro* transposon mutagenesis on *S. clavuligerus* DNA.

A predominant advantage of applying transposon mutagenesis is that disruption constructs can be prepared without having full sequence information for the region, and

therefore, a collection of transposon-disrupted clones can be used to sequence the fragment at the same time as individual disruption mutants are being prepared. In the present study, a partially sequenced cosmid, 6G9, which carries the cloned paralogue gene cluster and its flanking regions from *S. clavuligerus*, was subjected to the *in vitro* transposition reaction. By restriction digest analysis and Southern hybridization, the 11 kb *Eco*RI fragment from cosmid 6G9 was determined to be the one that lies next to *ceaS1* end of the paralogue gene cluster and therefore was chosen as the candidate for this study.

To locate the Tn*5* on the transposon-containing clones, DNA of transposon-containing clones was prepared and analyzed by restriction digestion followed by agarose gel electrophoresis. The Tn*5*-containing *Eco*RI fragment was determined by observing the disappearance of the original *Eco*RI fragment and the generation of new smaller fragments, due to the presence of two *Eco*RI restriction sites on the transposon. For those transposon-containing cosmids where it was particularly difficult to detect the change in size of *Eco*RI fragments, a *Hin*dIII and *Eco*RI double digestion was then used. This double digestion completely released the transposon from the target *Eco*RI fragment and therefore resulted in a more apparent change in size of the resulting fragments. However, due to the discrimination limitation of using conventional gel electrophoresis, this approach cannot provide 100% precision to the results, especially in the situations where transposon-insertion sites are very close to an *Eco*RI site or where the transposon inserts into a fragment larger than 10 kb. Therefore, this approach could only be used as a rough guide for transposon insertion mapping, and DNA sequencing was required to

determine transposon insertion sites accurately.

Although controlled by the "transposon immunity" mechanism, two transposon-containing cosmids were found to contain double Tn*5* insertions on the same target replicon. A double insertion was verified by the disappearance of two *Eco*RI fragments after electrophoresis. According to the protocol of the Tn*5* *in vitro* transposition kit (Epicentre), about 1% of the transposition clones may contain >1 transposon. Thus, the frequency of the occurrence of double Tn*5* insertions in this study, 2 out of 154 (1.3%), was close to the frequency reported in the kit. It should be clarified here that the two clones contain double Tn*5* insertions were not used for DNA sequencing analysis, since two primer-binding sites would undoubtedly give ambiguous sequencing results.

Tn*5* is the most random transposon system known. In the present study, this system retained the insertion characteristics of Tn*5* and exhibited little or no target site-specificity. For the 154 transposon-containing cosmids analyzed, a correlation has been observed between the frequencies of the transposon inserting into the different *Eco*RI fragments of cosmid 6G9, and the sizes of the *Eco*RI fragments. Further sequencing analysis of the subset of clones where the transposon had inserted into the 11 kb *Eco*RI fragment showed that the transposon was located at scattered sites on the fragment. After determining the precise insertion sites by DNA sequence analysis, the resulting data showed that Tn*5* transposase recognized a 9-bp sequence as its target, and a consensus target site sequence could be deduced: G-XCTCNNGGX-A (N = all bases, X = G or C). This consensus target site sequence not only matches one of the characteristics frequently found for Tn*5*

transposition target sites -- the occurrence of GC pairs at each end of the direct repeats (DR; Lodge et al. 1988), but also resembles the consensus target sequence suggested by Goryshin et al. at most DR positions except for position seven. The difference between the two consensus sequences may be attributed to the difference in G + C content between the two target DNAs, about 72% for *S. clavuligerus* in this study and about 50% for *E. coli* in Goryshin's study. All these results indicated that Tn*5* inserted quite randomly into *S. clavuligerus* DNA. Since the transposase is the only protein present in the *in vitro* transposition reaction, target selection should be a property of the transposase only. It could thus be concluded that the Tn*5* transposase functioned very well in the *in vitro* transposition system analyzed in this study.

After alignment of the DNA sequence information obtained by analysis of the transposon flanking regions, approximately 10 kb of contiguous DNA sequence that lacked *Eco*RI sites was obtained. This implies that transposon-containing cosmids that contain Tn*5* very close to the *Eco*RI sites (<1 kb) were not identified in this study. It is, again, mostly likely due to the limitation of discriminating the difference between 10 kb and 10-11 kb DNA fragments using the conventional gel electrophoresis. As a result, the transposon-containing cosmids that carry Tn*5* very close to the end of the target fragment were hard to identify. An improved approach able to discriminate subtle size differences among DNA fragments would be needed to find such mutant cosmids.

With the additional sequence needed to complete the 11kb *Eco*RI fragment being obtained by using specific oligonucleotide primers, the length of DNA fragment

sequenced in this study was 11443 bp, and the 11 kb *Eco*RI fragment on cosmid 6G9 was 11054 bp in length. Eleven open reading frames not previously described in *S. clavuligerus* were discovered as a result of the DNA sequencing. All of the new ORFs discovered were deduced from the DNA sequence by codon preference analysis and database homology searches, and thus these genes can only be assumed to produce their respective proteins. Both the average G + C content in the ORFs (72%) and the G +C frequencies in the third position of the ORF codons were in agreement with the characteristically biased codon preference of *Streptomyces* (Wright and Bibb 1992).

For six of the eleven ORFs, *Streptomyces* mutants were successfully constructed. In each case, a Tn*5* disrupted copy of the gene was crossed into the chromosome of wildtype *S. clavuligerus* by homologous recombination replacing the wildtype copy of the gene. Mutants were isolated based on their antibiotic resistance phenotype and then confirmed by Southern analysis. No aberrant growth phenotypes were found for all the mutants analyzed in this study. They were indistinguishable from the wildtype with respect to their ability to grow, differentiate and sporulate. This indicates that the genes disrupted play no apparent role in the growth and morphological development of *S. clavuligerus*. The involvement of each of the disrupted genes in the biosynthesis of clavams was then examined.

The putative protein encoded by *orfA* is a glycine/serine hydroxymethyltransferase. The *orfA* gene is located on the 2.8 kb *Eco*RI fragment adjacent to *ceaS1*, so it is the first gene beyond the left most end of the paralogue gene cluster. The HPLC analysis of the

"freeze-squeeze" supernatants showed that mutants with defects in *orfA* failed to produce alanylclavam but produced clavam-2-carboxylate, 2-hydroxymethylclavam and clavulanic acid. However, unlike mutants defective in either *orfC* or *orfD*, no new peak was generated. When analyzed in Soy liquid cultures, mutants defective in *orfA* produced various amounts of clavulanic acid and *5S* clavam metabolites. Bioassays of culture supernatants revealed that the mutants were unable to produce alanylclavam but produced wildtype levels of clavulanic acid and cephamycin C. Thus, OrfA may be involved in the biosynthesis of alanylclavam, but how this correlates with its similarities to glycine/serine hydroxymethyltransferases is unclear.

The *orfB* gene was identified when using primer walking to determine the sequence of the 5' end of the 11 kb *Eco*RI fragment of cosmid 6G9. Since no cosmid bearing the transposon in *orfB* was identified in this study, no mutant with defects in *orfB* has been constructed. The putative OrfB protein has strong homology to many hypothetical proteins. All these hypothetical proteins including OrfB contain a conserved domain from the YER057c/YjgF/UK114 family. In particular, OrfB has the strongest homology to the YjgF proteins. The YjgF protein belongs to a class of proteins of unknown function that exhibit striking conservation across a wide range of organisms, from bacteria to humans. Members of this family are identified by sequence similarity, and usually different names are given when present in different organisms.

The role for the OrfB protein in the biosynthesis of clavams is uncertain since its homologous proteins belong to a protein family of unknown function. However, recent

studies showed that a variety of biological processes seemed to be influenced by YjgF proteins. For example, the gene product of *aldR* from *Lactococcus lactis* has been implicated to be involved in isoleucine biosynthesis. In this case, the *aldR* mutant growth was slowed down twofold in the absence of isoleucine, suggesting that AldR may interfere with the synthesis of isoleucine. However, AldR is not an enzyme directly involved in isoleucine biosynthesis. Further studies are needed to investigate its putative function (Goupil-Feuillerat et al. 1997). In another example, characteristics of mutants defective in *yjgF* from *S. typhimurium* suggested that the YjgF protein, or alternatively, the product of the YjgF protein if it functions as an enzyme, is important for appropriate function and/or regulation of the isoleucine biosynthetic pathway. Furthermore, carbon flux through the isoleucine biosynthetic pathway is important for an *yjgF* mutation to restore thiamine synthesis (Enos-Berlage et al. 1998). In the purine operon of *B. subtilis*, adenine mediates a PurR-dependent repression of *purA* transcription. PurR is a purine repressor and the *purA* gene encodes adenylosuccinate synthetase. However, the repression by PurR is dependent upon YabJ, an YjgF family protein, by an unknown mechanism (Rappu et al. 1999).

Furthermore, the role of YjgF homologous proteins in Eukaryotes is more diverse. Homologs from rat and human have been shown to inhibit cell-free protein synthesis when present at a high concentration (Oka et al. 1995). The YabJ homolog Hrp12 from mouse has some similarity to heat shock proteins Hsp70 and Hsp90, and the purified mouse Hrp12 could be phosphorylated *in vitro* with protein kinase C (Samuel et al. 1997).

Therefore, no certain biological function or common molecular function has emerged from the studies of YjgF proteins. Until mutants of the *orfB* gene have been generated by gene replacement and their ability to synthesize clavams has been analyzed by HPLC, it is hard to predict whether the *orfB* gene plays a role in β–lactam biosynthesis.

The putative OrfC and OrfD proteins likely act as a PLP-dependent aspartate aminotransferase and threonine dehydratase, respectively. Aspartate aminotransferase is an enzyme that catalyzes the reversible transfer of an amino group from aspartate to α-ketoglutarate to form glutamate and oxaloacetate, while threonine dehydratase catalyzes the dehydration of threonine (and serine) to yield 2-ketobutyrate (and pyruvate) both using PLP as a cofactor.

The mutants defective in either *orfC* or *orfD* were analyzed for their ability to synthesize clavams by fermentation and bioassays. It should be noted that *5S* clavam production is highly sensitive to growth conditions and growth stages. It has been shown that *5S* clavam production tends to be influenced very easily by agitation and most probably by aeration. More *5S* clavam metabolites were produced when foil was used to seal the opening of culture-containing flasks during fermentation (Jensen et al., personal communications), and moreover, cultures tended to produce more *5S* clavam metabolites when grown on Soy solid medium but more clavulanic acid when grown on Soy liquid medium. In addition, considerable variation in *5S* clavam production between replicate cultures of the same strains was very often observed. Thai et al. reported that even though DNA measurements suggested that the extents of growth in replicate cultures was very

similar, variation in *5S* clavam production was still observed (Thai et al. 2001). Therefore, only dramatic and consistent changes in *5S* clavam production in this study were concluded to be of significance.

An HPLC analysis of the "freeze-squeeze" supernatants obtained from patches of cultures grown on solid Soy medium showed that mutants with defects in either *orfC* or *orfD* failed to produce alanylclavam but produced clavam-2-carboxylate, 2-hydroxymethylclavam and clavulanic acid. However, a new peak with unknown composition was also generated when these mutants were cultivated on Soy solid medium. When the mutants were analyzed in Soy liquid cultures, both mutant types produced various amounts of clavam-2-carboxylate, 2-hydroxymethylclavam and clavulanic acid, but mutants that are defective in *orfD* tended to produce lesser amounts of all of the clavams in general. However, no new peak could be detected when the mutants were cultivated in Soy liquid media, suggesting that the new clavam might only be produced/accumulated to a detectable level on Soy solid medium. Likewise, bioassays of culture supernatants revealed that both mutant types were unable to produce alanylclavam but produced wildtype levels of clavulanic acid and cephamycin C. Thus, OrfC and OrfD may be involved in the biosynthesis of alanylclavam, but how this correlates with their similarities to aspartate aminotransferases and threonine dehydratases is unclear.

Previous studies on the biosynthesis of tylosin in *S. fradiae* provide an example where amino acid catabolism is a vital source of certain antibiotic precursors in actinomycetes. In this study, aspartate aminotransferase as well as valine dehydrogenase

and threonine dehydratase was required for the biosynthesis of tylosin in *S. fradiae* (Lee and Lee 1991). The change in specific tylosin production rates in batch cultures showed patterns similar to those of the specific production rates of aspartate aminotransferase, valine dehydrogenase, and threonine dehydratase (Lee and Lee 1993; Lee and Lee 1994). Further analysis showed that targeted inactivation of the valine dehydrogenase gene (*vdh*) generated mutants that produced much lower level of tylosin (Tang et al. 1994).

When it comes to the present study, the OrfC and OrfD proteins may function in the biosynthesis of alanylclavam in *S. clavuligerus* by providing amino acid catabolites as necessary precursors. In the tentative pathway to *5S* clavam biosynthesis in *S. clavuligerus* proposed by Egan et al., it was suggested that clavaminic acid is successively decarboxylated and deaminated by way of a PLP-dependent enzyme(s) to produce a branch-point intermediate that is then converted to *5S* clavams (Egan et al. 1997). However, since mutants defective in *orfC* are only affected the biosynthesis of alanylclavam, it is therefore unlikely that the putative PLP-dependent animotransferase identified in this study is the candidate enzyme to produce such a branch-point intermediate leading to various *5S* clavams. On the other hand, the OrfC and OrfD are more likely to be examples of enzymes that catalyze the biosynthesis of particular clavams that we have been looking for, and in this case, they may catalyze the biosynthesis of alanylclavam in *S. clavuligerus*. Further analysis of these two proteins is required before a firm conclusion can be reached.

The putative OrfE protein contains a stretch of hydrophobic amino acids present at

the N-terminus of the protein, suggesting OrfE is membrane-anchored. The role of OrfE could not be deduced since the protein does not show significant homology to any known proteins during database searches nor could any conserved domains be found. The deduced OrfF protein contains a number of transmembrane domains capable of inserting the 25.7-KD protein into the cytoplasmic membrane. The protein shows good similarities to several membrane proteins and contains a conserved domain characteristic of predicted AzlC branched-chain amino acid permease for azaleucine resistance. The AzlC branched-chain amino acid permease usually consists of pairs of integral membane proteins like the AzlC and AzlD proteins from *B. subtilis*, which comprise an efflux pump for branched chain amino acids, leucine, isoleucine and valine (Belitsky and Gustafsson 1997). However, the OrfE protein does not resemble AzlD of *B. subtilis* in amino acid sequence and lacks the conserved domain characteristic of AzlD branched-chain amino acid permeases. Thus the newly discovered OrfE and OrfF proteins of *S. clavuligerus* seems unlikely to function like the AzlC and AzlD proteins of *B. subtilis*. Further database searches showed that the two homologous proteins of OrfE and OrfF from *S. coelicolor*, the hypothetical protein SC4H8.03 and the putative membrane protein SC4H8.02, also lie next to each other on the chromosome of *S. coelicolor*. It thus seems unlikely that the genes encoding these two membrane proteins should be located together in two *Streptomyces* spp. just by chance. It is possible that the OrfE and OrfF proteins function together towards a membrane associated common goal.

The deduced OrfG protein shows a high degree of similarity to a small number of

hypothetical proteins from *Streptomyces* spp. in the databases, and no conserved domains could be detected in OrfG. Therefore the function of the *orfG* gene in *S. clavuligerus* is unknown. The deduced OrfH protein contains a number of transmembrane domains capable of inserting the protein into the cytoplasmic membrane. OrfH shows a moderate degree of similarity to many membrane proteins and contains a conserved domain characteristic of AraJ arabinose efflux permease. AraJ belongs to a large class of multidrug resistance translocators, and in particular to the major facilitator superfamily (MFS). MFS is the main class of proteins that utilize protons as symporters or antiporters to transport a variety of molecules across the cytoplasmic membrane of bacteria and eukaryotes. The specificity of these proteins is very broad, and their substrates form a diverse group of compounds including antibiotics and sugars.

Although three separate attempts were made, genes *orfE*, *orfF*, *orfG* and *orfH* were unable to be disrupted in the wildtype *S. clavuligerus* background. This might be due to technical problems of applying conjugation to *S. clavuligerus*. It is known that similar conjugation technology usually generates 100-300 exconjugants on *S. coelicolor*, but far fewer exconjugants, usually 5-30, can be obtained on *S. clavuligerus* even though the same numbers of *Streptomyces* spores are used. Further improvements on conjugation technology are needed in order to increase the conjugation frequency of *S. clavuligerus*.

On the other hand, since not a single exconjugant could be obtained after repetitions of the experiments, it is possible that *orfE*, *orfF*, *orfG* and *orfH* may be essential genes, and disruption of these genes would be lethal to the bacterial cells. Recently, it has been

estimated that the *M. tuberculosis* genome contains 35% essential genes. In that study, with the knowledge of all sites on the *M. tuberculosis* genome eligible for transposon mutagenesis and an experimentally determined partial list of nonessential genes from genome mutagenesis, a Bayesian statistical method was used to predict genes that are likely to be essential and estimate their proportion in haploid genomes. This analysis further identified several classes of genes with high probabilities of being enriched in essential genes, and probable integral membrane proteins and conserved hypothetical proteins are among the first four classes of genes with the highest probabilities of being essential (Lamichhane et al. 2003). Mycobacteria and streptomycetes are both actinomycetes sharing a high G + C content, and it is known that fast-growing mycobacteria and *Streptomyces* share a similar habitat in the soil. At the molecular level, studies showed that mycobacterial genes are expressed from their own promoters in *Streptomyces* (Kieser et al. 1986). Therefore, the essential-gene prediction results determined for mycobacteria could be applicable to *Streptomyces* genes as well.

In this study, four out of eleven (36%) genes disrupted seem to be essential. Although this percentage is not statistically significant since only a very small number of genes were analyzed, it is still very close to the percentage of essential genes estimated in *M. tuberculosis* genome. Furthermore, the deduced OrfF and OrfH proteins both possess the characteristics of integral membrane proteins, which are among the gene groups with the highest probabilities of being essential, as was predicted in *M. tuberculosis*. Since another membrane-anchored protein, OrfE, is predicted to function with OrfF towards a

membrane associated common goal, disruption in *orfE* may destroy the function of the OrfE/OrfF pair of proteins and thus is also likely to cause a lethal mutation. On the other hand, OrfH protein likely acts as a transporter that exports antibiotics or sugars from the cell. It thus may serve to protect the cell from the potential lethal effects of the metabolites that it produces. For the OrfG protein, although no putative function could be predicted, its possibility of being essential could not be ruled out since very little is known about the function of conserved hypothetical proteins. In addition, according to Lamichhane et al., conserved hypothetical proteins are among the gene groups with the highest probability of being essential.

Another possible essential function for OrfG can be explained in the following way: the OrfG protein shows significant similarity to two hypothetical proteins from *S. coelicolor* and *S. avermitilis*. Although the function of these hypothetical proteins is unknown, genes of unknown function that are conserved between organisms are more likely to be essential than non-conserved genes. In addition, unlike *S. clavuligerus*, neither of these *Streptomyces* spp. produces clavams. Therefore, OrfG may play some role in both primary and secondary metabolism, such as amino acid metabolism, and thus be essential for cell survival but also involved in clavam production in *S. clavuligerus*.

In order to test whether the *orfE*, *orfF*, *orfG* and *orfH* genes are essential in *S. clavuligerus*, a conventional way to approach the problem would be the generation of conditional mutations within the genes that affect growth. Here construction of a complementing plasmid that carries the mutant version of the gene is required. The

mutation in this second copy of the gene is usually conditional, which means the expression of the mutant gene is, for instance, temperature sensitive or inducible. After generation of gene disruption of the chromosomal copy by homologous recombination, the essentiality of the gene could be identified by growth under a permissive condition but not a non-permissive condition.

For the convenience of analyzing essential genes that may be encountered in future studies during transposon mutagenesis, the transposon could be modified to contain an outward-facing inducible promoter at one end of the transposon. In this way, transposition with such a transposon into the promoter region of a gene can create a gene in which an inducible promoter replaces the function of the natural promoter. If the gene is essential for survival, the bacteria are now dependent on the inducer for growth. An example of the application of this approach to define genes essential for growth is reported by Judson and Mekalanos. They successfully used an arabinose-inducible promoter in a transposon termed TnAraOut to define 16 insertions that showed an arabinose-dependent growth phenotype. Five of these insertions defined genes essential for colony formation on LB, and 11 defined ORFs that showed a reduced growth rate in the absence of arabinose. The latter insertions defined the genes that are not essential but whose expression is required for optimum growth (Judson and Mekalanos 2000). This approach represents one of the potential improvements that could be made to broaden the application of transposon mutagenesis.

The Putative Orf1 protein shows high similarity to transcriptional regulatory proteins

of the LysR family. As occurs with all of the transcriptional regulators of the LysR family, OrfI also contains a putative DNA-binding HTH motif at the N-terminus. Mutants with defects in *orfI* were still able to produce alanylclavam, clavam-2-carboxylate, 2-hydroxymethylclavam and clavulanic acid on Soy solid medium. When grown on Soy liquid medium, mutants produced various amounts of clavam-2-carboxylate, 2-hydroxymethylclavam and clavulanic acid. Bioassays of culture supernatants revealed that mutants produced reduced levels of alanylclavam but wildtype levels of clavulanic acid and cephamycin C.

Simply based on the fermentation and bioassay results, it is hard to predict the involvement of the *orfI* gene in the biosynthesis of clavams in *S. clavuligerus* since clavam production is very variable from batch to batch during different fermentations. However, we cannot eliminate the possibility either, because clustering of antibiotic biosynthesis gene with regulatory genes that control a specific pathway is probably a general phenomenon in *Streptomyces*. A SARP type transcriptional regulator, CcaR, which is present in the cephamycin gene cluster, is found to co-regulate the biosynthesis of cephamycin C, clavulanic acid and *5S* clavam metabolites (Perez-Llarena et al. 1997, Alexander and Jensen 1998, Jensen et al. unpublished data). A LysR-like protein, ClaR, which specifically targets the regulation of clavulanic acid, is encoded by the *claR* gene present in the clavulanic acid gene cluster. ClaR targets particularly the "late" biosynthetic genes involved in the steps subsequent to clavaminic acid formation (Paradkar et al. 1998). During our studies of the genes flanking *cas1*, the deduced product

of a newly identified gene, *cvm7*, shows a high degree of similarity to SARPs of other antibiotic producers including similarity to CcaR. It was then predicted that this *cvm*-SARP might be a pathway specific regulator controlling production some or all of the *5S* clavam genes. However, disruption of *cvm7* did not result in loss of *5S* clavam production and thus its involvement in clavam biosynthesis is unclear (Jensen et al. unpublished data). Till now, very little is known about the regulation of the genes of *5S* clavam metabolites. Therefore, characterization of *orfI* is of great interest.

The putative OrfJ protein showed significant similarities to many hypothetical proteins. They share a conserved domain characteristic of FolA dihydrofolate reductase for coenzyme metabolism. The putative OrfK protein shows low similarities to a limited number of hypothetical proteins, and they share a conserved domain of unknown function. The putative OrfL protein shows similarities to transcriptional regulators of ArsR family. The putative OrfL protein and its homologous proteins share a conserved domain characteristic of arsenical resistance operon repressor and similar prokaryotic, metal-regulated homodimeric repressors. However, no putative open reading frames that resemble those present in the *ars* operon could be identified in the *orfL* flanking regions. Therefore, the putative OrfL protein is unlikely to belong to an *ars* operon, but still may possess similar regulatory functions in response to arsenite or other metal ions.

An HPLC analysis of the "freeze-squeeze" supernatants showed that mutants that are defective in *orfJ*, *orfK* or *orfL* were able to produce alanylclavam, clavam-2-carboxylate, 2-hydroxymethylclavam and clavulanic acid on Soy solid medium. When analyzed in

Soy liquid cultures, all mutant types produced various amounts of clavam-2-carboxylate, 2-hydroxymethylclavam and clavulanic acid. Bioassays of culture supernatants showed that all the mutant types were able to produce alanylclavam, clavulanic acid and cephamycin C. Thus, mutants disrupted in *orfJ*, *orfK* or *orfL* did not lose their ability to synthesize clavams in *S. clavuligerus*.

However, recent identification by Southern hybridization of additional paralogue genes for some of the *cvm* genes is making the story more complicated (Jensen et al. unpublished data). If these genes are true paralogues, disruption of only one copy of the putative clavam metabolite biosynthetic genes may not completely block the production of *5S* clavam and thus will not result in a dramatic phenotype, unless the second copy of the gene can be identified and knocked out. It is possible that as yet undiscovered paralogues also exist for the genes identified in this study. These paralogues might be able to compensate for the loss of part of the clavam biosynthetic ability of the mutants. Southern hybridization could be used to identify putative paralogues for these genes.

A transposon-generated mutation in an upstream gene of a co-transcriptional unit will exert polar effects on the downstream gene. For instance, the *orfC* and *orfD* genes seem likely to be transcriptionally coupled. Although disruption of either *orfC* or *orfD* genes generated similar clavam production patterns, if polar effects appear to be an issue, disruption of the upstream gene *orfC* could result in the interference of the transcription of *orfD* as well. Therefore, in the situation when either the *orfC* or *orfD* gene was disrupted, the transcription of *orfD* would be blocked. Thus it is possible that only the

orfD gene is responsible for loss of the ability to produce alanlyclavam rather than both *orfC* and *orfD*. To further analyze the involvement of each gene of a co-transcriptional unit in clavam production, mutants can then be subjected to a second round of conventional gene replacement mutation to excise the transposon, leaving in its place a less disruptive deletion or frameshift mutation.

In summary, this study represents the first example of successful application of the Tn*5*-based *in vitro* transposon mutagenesis system on *S. clavuligerus*. As one of the main advantages of this technology over the conventional gene disruption approaches, disruption constructs were prepared without having full sequence information for the region. Transposon-disrupted clones were screened by restriction digestion and a collection of mutants was used to sequence the fragment at the same time as individual disruption mutants were prepared. As a result, sequence of an 11.4 kb DNA fragment, flanking one end of the newly discovered paralogue gene cluster, was determined, and 11 putative ORFs were identified in this region. Mutant clones were introduced into wildtype *S. clavuligerus* via conjugation to create mutant strains by gene replacement selecting for double-crossover. Antibiotic biosynthesis of the mutant strains was detected by bioassay and HPLC. However, although proposed gene products could be deduced by homology searches from the database and their potential involvements in clavam biosynthesis were analyzed by bioassay and HPLC, the exact role and significance of the deduced products are uncertain mainly because little is known about the biochemistry of the pathway leading from clavaminic acid to the various clavam metabolites produced by

S. clavuligerus, and thus further studies are needed.

In addition to the possible future work mentioned in this section, other future work is reqiured on many aspects of the newly identified genes. For example, the fermentation for *5S* clavam production was difficult in that it did not provide easily reproducible results. Ideally, additional work should be done to optimize fermentation conditions, especially the availability of oxygen, agitation, inoculum size, etc. Improvement of conjugation technology is also needed to increase the conjugation frequency. Disruption of *orfB* would be of great interest. The *orfB* gene is located between *orfA* and *orfC*. Since mutants in either *orfA* or *orfC* both lacked the ability to produce alanylclavam, and antibiotic genes are always clustered on the chromosome of bacteria, it is possible that *orfB* is also involved in the clavam biosynthetic pathway. With the availability of sequence information, disruption of *orfB* can be done by conventional gene disruption methods such as insertion of an antibiotic resistance marker into a restriction site within the gene. For genes encoding putative enzymes, the biochemical function of the gene product can be assayed. The regulation target(s) for OrfI, the LysR family transcriptional regulator can be investigated by Northern hybridization. The promoter-less *egfp* gene carried by the transposon will facilitate the *in vivo* analysis of the disrupted genes. For example, use of EGFP as a reporter can be helpful for *in vivo* localization of the disrupted protein, and transcriptional analysis of disrupted genes at different time point of cell development and under different growth conditions.

V. REFERENCES

Aharonowitz, Y., and A.L. Demain. 1978. Carbon catabolite regulation of cephamycin production in *Streptomyces clavuligerus*. Antimicrob. Agents Chemother. 14:159-164.

Alexander, D.C., and S.E. Jensen. 1998. Investigation of the *Streptomyces clavuligerus* cephamycin C gene cluster and its regulation by the CcaR protein. J. Bacteriol. 180:4068-4079.

Bachmann, B.O., R. Li, and C.A. Townsend. 1998. *β*-Lactam synthetase: a new biosynthetic enzyme. Proc. Natl. Acad. Sci. U.S.A. 95:9082-9086.

Bachmann, B.O., and C.A. Townsend. 2000. Kinetic mechanism of the *β*-lactam synthetase of *Streptomyces clavuligerus*. Biochemistry 39:11187-11193.

Baggaley, K.H., A.G. Brown, and C.J. Schofield. 1997. Chemistry and biosynthesis of clavulanic acid and other clavams. Nat. Prod. Rep. 14:309-333.

Baltz, R.H., M.A. McHenney, C.A. Cantwell, S.W. Queener, and P.J. Solenberg. 1997. Application of transposition mutagenesis in antibiotic producing *Streptomyces*. Antonie van Leeuvwenhoek 71:179-187.

Belitsky, B.R., M.C.U. Gustafsson, A.L. Sonenshein, and C.V. Wachenfeldt. 1997. An *lrp*-like gene of *Bacillus subtilis* involved in branched-chain amino acid transport. J. Bacteriol. 179:5448-5457.

Bender, J., and N. Kleckner. 1992. IS*10* transposase mutations that specifically alter target site recognition. EMBO J. 11:741-750.

Berg, D.E. 1989. Transposon Tn*5*. p. 185-210. In D.E. Berg and M.M. Howe. (eds.) Mobile DNA. American Society for Microbiology. Washington D.C.

Berg, C.M., D.E. Berg, and E.A. Groisman. 1989. Transposable elements and the genetic engineering of bacteria. p. 879-925. In D.E. Berg and M.M. Howe. (eds.) Mobile DNA. American Society for Microbiology. Washington D.C.

Chater, K.F., C.J. Bruton, S.G. Foster, and I. Tobek. 1985. Physical and genetic analysis of *IS110*, a transposable element of *Streptomyces coelicolor* A3(2). Mol. Gen. Genet. 200:235-239.

Chung, S.T. 1987. Tn*4556*, a 6.8-kilobase-pair transposable element of *Streptomyces*

fradiae. J. Bacteriol. 169:4436-4441.

Davis, N.K., and K.F. Chater. 1990. Spore colour in *Streptomyces coelicolor* A3(2) involves the developmentally regulated synthesis of a compound biosynthetically related to polyketide antibiotics. Mol. Microbiol. 4:1679-1691.

de La Fuente, A., L.M. Lorenzana, J.F. Martin, and P. Liras. 2002. Mutants of *Streptomyces clavuligerus* with disruptions in different genes for clavulanic acid biosynthesis produce large amounts of holomycin: possible cross-regulation of two unrelated secondary metabolic pathways. J. Bacteriol. 184:6559-6565.

De Lorenzo, V., L. Eltis, B. Kessler, and K.N. Timmis. 1993. Analysis of *Pseudomonas* gene production using *lacI*^{q}/P*trp-lac* plasmids and transposons that confer conditional phenotypes. Gene. 123:17-24.

DeBoy, R.T., and N.L. Craig. 2000. Target site selection by Tn*7*: *att*Tn*7* transcription and target activity. J. Bacteriol. 182:3310-3313.

Demain, A.L., and R.P. Elander. 1999. The β-lactam antibiotics: past, present, and future. Antonie Van Leeuwenhoek 75:5-19.

Douglas, S.E. 1995. DNA Strider. An inexpensive sequence analysis package for the Macintosh. Mol. Biotechnol. 3:37-45.

Egan, L.A., R.W. Busby, D. Iwata-Reuyl, and C.A. Townsend. 1997. Probable role of clavaminic acid as the terminal intermediate in the common pathway to clavulanic acid and the *5S* clavam metabolites. J. Am. Chem. Soc. 119:2348-2355.

Elson, S.W., K.H. Baggaley, M. Davison, M. Fulston, N.H. Nicholson, G.D. Risbridger, and J.W. Tyler. 1993. The identification of three new biosynthetic intermediates and one further biosynthetic enzyme in the clavulanic acid pathway. J. Chem. Soc., Chem. Commun. 1993:1212-1214.

Elson, S.W., K.H. Baggaley, J. Gillett, S. Holland, N.H. Nicholson, J.T. Sime, and Woroniecki S.R. 1987. Isolation of two novel intracellular β-lactams and a novel dioxygenase cyclising enzyme from *Streptomyces clavuligerus*. J. Chem. Soc., Chem. Commun. 1987:1736-1738.

Engel, P., and A.R. Lax. 1997. Partial deletion of transposon Tn*4560* integrated into the genome of *Streptomyces tendae*. Lett. Appl. Microbiol. 25:225-228.

Engel, P., L.L. Scharfenstein, J.M. Dyer, and J.W. Cary. 2001. Disruption of a gene encoding a putative gamma-butyrolactone-binding protein in *Streptomyces tendae* affects nikkomycin production. Appl. Microbiol. Biotechnol. 56:414-419.

Engel, P., and M.S. Wright. 1993. Use of Tn*4560* to generate nikkomycin non-producing mutants of *Streptomyces tendae*. FEMS Microbiol. Lett. 109:257-261.

Enos-Berlage, J.L., M.J. Langendorf, and D.M. Downs. 1998. Complex metabolic phenotypes caused by a mutation in *yjgF*, encoding a member of the highly conserved YER057c/YjgF family of proteins. J. Bacteriol. 180:6519-6528.

Fernandez-Moreno, M.A., J.L. Caballero, D.A. Hopwood, and F. Malpartida. 1991. The *act* cluster contains regulatory and antibiotic export genes, direct targets for translation control by the *bldA* transfer RNA gene of *Streptomyces*. Cell 27:769-780.

Gagnat, J., H. Chouayekh, C. Gerbaud, F. Francou, and M.J. Virolle. 1999. Disruption of *sblA* in *Streptomyces lividans* permits expression of a heterologous alpha-amylase gene in the presence of glucose. Microbiology 145:2303-2312.

Gehring, A.M., J.R. Nodwell, S.M. Beverley, and R. Losick. 2000. Genomewide insertional mutagenesis in *Streptomyces coelicolor* reveals additional genes involved in morphological differentiation. Proc. Natl. Acad. Sci. U.S.A 97:9642-9647.

Goryshin, I.Y., J. Jendrisak, L.M. Hoffman, R. Meis, and W.S. Reznikoff. 2000. Insertional transposon mutagenesis by electroporation of released Tn*5* transposition complexes. Nat. Biotechnol. 18:97-100.

Goryshin, I.Y., J.A. Miller, Y.V. Kil, V.A. Lanzov, and W.S. Reznikoff. 1998. Tn*5*/IS*50* target recognition. Proc. Natl. Acad. Sci. U.S.A. 95:10716-10721.

Goupil-Feuillerat, N., M. Cocaign-Bousquet, J.J. Godon, S.D. Ehrlich, and P. Renault. 1997. Dual role of alpha-acetolactate decarboxylase in *Lactococcus lactis* subsp. Lactis. J. Bacteriol. 179:6285-6293.

Hanahan, D. 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. 166:557-580.

Haren, L., B. Ton-Hoang, and M. Chandler. 1999. Integrating DNA: transposases and retroviral integrases. Annu. Rev. Microbiol. 53:245-281.

Hedges, R.W., and A.E. Jacob. 1974. Transposition of Amp resistance from RP4 to other

replicons. Mol. Gen. Genet. 132:31-40.

Herron, P.R., M.C. Evans, and P.J. Dyson. 1999. Low target site specificity of an IS*6100*-based mini-transposon, Tn*1792*, developed for transposon mutagenesis of antibiotic-producing *Streptomyces*. FEMS Microbiol. Lett. 171:215-221.

Hodgson, J.E., A.P. Fosberry, N.S. Rawlinson, H.N. Ross, R.J. Neal, J.C. Arnell, A.J. Earl, and E.J. Lawlor. 1995. Clavulanic acid biosynthesis in *Streptomyces clavuligerus*: gene cloning and characterization. Gene. 166:49-55.

Ichikawa, H., K. Ikeda, J. Amemura, and E. Ohtsubo. 1990. Two domains in the terminal inverted-repeat sequence of transposon Tn*3*. Gene 86:11-17.

Ikeda, H., Y. Takada, C.H. Pang, H. Tanaka, and S. Omura. 1993. Transposon mutagenesis by Tn*4560* and applications with avermectin-producing *Streptomyces avermitilis*. J. Bacteriol. 175:2077-2082.

Ishikawa, J., and K. Hotta. 1999. FramePlot: a new implementation of the frame analysis for predicting protein-coding regions in bacterial DNA with a high G + C content. FEMS Microbiol. Lett. 174:251-253.

Jensen, R.A., and W. Gu. 1996. Evolutionary recruitment of biochemically specialized subdivisions of Family I within the protein superfamily of aminotransferases. J. Bacteriol. 178:2161-2171.

Jensen, S.E., K.J. Elder, K.A. Aidoo, and A.S. Paradkar. 2000. Enzymes catalyzing the early steps of clavulanic acid biosynthesis are encoded by two sets of paralogue genes in *Streptomyces clavuligerus*. Antimicrob. Agents Chemother. 44:720-726.

Jensen, S.E., A.S. Paradkad, R.H. Mosher, C. Anders, P.H. Beatty, M.J. Brumlik, A. Griffin, and B. Barton. 2003a. Five additional genes are involved in clavulanic acid biosynthesis in *Streptomyces clavuligerus*. Manuscripts.

Jensen, S.E., A. Wong, A. Griffin, and B. Barton. 2003b. *Streptomyces clavuligerus* has a second copy of the gene encoding proclavaminate amidinohydrolase. Manuscript.

Judson, N., and J.J. Mekalanos. 2000. TnAraOut, a transposon-based approach to identify and characterize essential bacterial genes. Nat. Biotechnol. 18:740-745.

Kershaw, N.J., H.J. McNaughton, K.S. Hewitson, H. Hernandez, J. Griffin, C. Hughes, P. Greaves, B. Barton, C.V. Robinson, and C.J. Schofield. 2002. ORF6 from the clavulanic

acid gene cluster of *Streptomyces clavuligerus* has ornithine acetyltransferase activity. Eur. J. Biochem. 269:2052-2059.

Khaleeli, N., R. Li, and C.A. Townsend. 1999. Origin of the β-lactam carbons in clavulanic acid from a usual thiamine pyrophosphate-mediated reaction. J. Am. Chem. Soc. 121:9223-9224.

Kieser, T., M.J. Bibb, M.J. Buttner, K.F. Chater, and D.A. Hopwood. 2000. General introduction to actinomycete biology. p. 2-21. Practical *Streptomyces* Genetics. The John Innes Foundation, Norwich, U.K.

Kyte, J., and R. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157:105-132.

Lamichhane, G., M. Zignol, N.J. Blades, D.E. Geiman, A. Dougherty, J. Grosset, K.W. Broman, and W.R. Bishai. 2003. A postgenomic method for predicting essential genes at subsaturation levels of mutagenesis: application to *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. U.S.A. 100:7213-7218.

Lee, S.H., and K.J. Lee. 1991. Relationship between threonine dehydratase and biosynthesis of tylosin in *Streptomyces fradiae*. J. Gen. Microbiol. 137:2547-2553.

Lee, S.H., and K.J. Lee. 1993. Aspartate aminotransferase and tylosin biosynthesis in *Streptomyces fradiae*. Appl. Environ. Microbiol. 59:822-827.

Lee, S.H., and K.J. Lee. 1994. Kinetics of the repression of tylosin biosynthesis by ammonium ion in *Streptomyces fradiae*. J. Biotechnol. 32:149-156.

Li, R., N. Khaleeli, and C.A. Townsend. 2000. Expansion of the clavulanic acid gene cluster: identification and *in vivo* functional analysis of three new genes required for biosynthesis of clavulanic acid by *Streptomyces clavuligerus*. J. Bacteriol. 182:4087-4095.

Liras, P. 1999. Biosynthesis and molecular genetics of cephamycins. Cephamycins produced by actinomycetes. Antonie Van Leeuwenhoek 75:109-124.

Lodge, J.K., K. Weston-Hafer, and D.E. Berg. 1988. Transposon Tn*5* target specificity: preference for insertion at G/C pairs. Genetics. 120:645-650.

MacNeil, D.J. 1988. Characterization of a unique methyl-specific restriction system in *Streptomyces avermitilis*. J. Bacteriol. 170:5607-5612.

Marsh, E.N., M.D.T. Chang, and C.A. Townsend. 1992. Two isoenzymes of clavaminate synthase central to clavulanic acid formation: cloning and sequencing of both genes from *Streptomyces clavuligerus*. Biochemistry 31:12648-12657.

Martin, J.F. 1998. New aspects of genes and enzymes for beta-lactam antibiotic biosynthesis. Appl. Microbiol. Biotechnol. 50:1-15.

Mehta, P.K., T.I. Hale, and P. Christen. 1993. Aminotransferases: demonstration of homology and division into evolutionary subgroups. Eur. J. Biochem. 214:549-561.

Mellado, E., L.M. Lorenzana, M. Rodriguez-Saiz, B. Diez, P. Liras, and J.L. Barredo. 2002. The clavulanic acid biosynthetic cluster of *Streptomyces clavuligerus*: genetic organization of the region upstream of the *car* gene. Microbiology 148:1427-1438.

Mosher, R.H., A.S. Paradkar, C. Anders, B. Barton, and S.E. Jensen. 1999. Genes specific for the biosynthesis of clavam metabolites *5S* to clavulanic acid are clustered with the gene for clavaminate synthase 1 in *Streptomyces clavuligerus*. Antimicrob. Agents Chemother. 43:1215-1224.

Nagarajan, R., L.D. Boeck, M. Gorman, R.L. Hamill, C.E. Higgens, M.M. Hoehn, W.M Stark, and J.G. Whitney. 1971. Beta-lactam antibiotics from *Streptomyces*. J. Am. Chem. Soc. 93:2308-2310.

Nathisuwan, S., D.S. Burgess, and J.S. Lewis 2nd. 2001. Extended-spectrum beta-lactamases: epidemiology, detection, and treatment. Pharmacotherapy 21:920-928.

Neyfakh, A.A., V.E. Bidnenko, and L.B. Chen. 1991. Efflux-mediated multidrug resistance in *Bacillus subtilis*: similarities and dissimilarities with the mammalian system. Proc. Natl. Acad. Sci. U.SA. 88:4781–4785.

Nicholson, N.H., K.H. Baggaley, R. Cassels, M. Davison, S.W., Elson, M. Fulston, J.W. Tyler, and S.R. Woroniecki. 1994. Evidence that the immediate biosynthetic precursor of clavulanic acid is its *N*-aldehyde analogue. J. Chem. Soc. Chem. Commun. 1994:1281-1282.

Oka, T., H. Tsuji, C. Noda, K. Sakai, Y. Hong, I. Suzuki, S. Munoz, and Y. Natori. 1995. Isolation and characterization of a novel perchloric acid-soluble protein inhibiting cell-free protein synthesis. J. Biol. Chem. 270:30060-30067.

Paradkar, A.S., K.A. Aidoo, and S.E. Jensen. 1998. A pathway-specific transcriptional activator regulates late steps of clavulanic acid biosynthesis in *Streptomyces clavuligerus*.

Mol. Microbiol. 27:831-843.

Paradkar, A.S., and S.E. Jensen. 1995. Functional analysis of the gene encoding the clavaminate synthase 2 isoenzyme involved in clavulanic acid biosynthesis in *Streptomyces clavuligerus*. J. Bacteriol. 177:1307-1314.

Paranthaman, S., and K. Dharmalingam. 2003. Intergeneric conjugation in *Streptomyces peucetius* and *Streptomyces* sp. strain C5: chromosomal integration and expression of recombinant plasmids carrying the *chiC* gene. Appl. Environ. Microbiol. 69:84-91.

Perez-Llarena, F.J., P. Liras, A. Rodriguez-Garcia, and J.F Martin. 1997. A regulatory gene (*ccaR*) required for cephamycin and clavulanic acid production in *Streptomyces clavuligerus*: amplification results in overproduction of both beta-lactam compounds. J. Bacteriol. 179:2053-2059.

Perez-Redondo, R., A. Rodriguez-Garcia, J.F. Martin, and P. Liras. 1998. The *claR* gene of *Streptomyces clavuligerus*, encoding a LysR-type regulatory protein controlling clavulanic acid biosynthesis, is linked to the clavulanate-9-aldehyde reductase (*car*) gene. Gene 12:311-321.

Perez-Redondo, R., A. Rodriguez-Garcia, J.F. Martin, and P. Liras. 1999. Deletion of the *pyc* gene blocks clavulanic acid biosynthesis except in glycerol-containing medium: evidence for two different genes in formation of the C3 unit. J. Bacteriol. 181:6922-6928.

Pitman, A., P. Herron, and P. Dyson. 2002. Cointegrate resolution following transposition of Tn*1792* in *Streptomyces avermitilis* facilitates analysis of transposon-tagged genes. J. Microbiol. Methods 49:89-96.

Pribil, P.A., and D. Haniford. 2000. Substrate recognition and induced DNA deformation by transposase at the target-capture stage of Tn*10* transposition. J. Mol. Biol. 303:145-159.

Pruess, D.L., and M. Kellett. 1983. Ro 22-5417, a new clavam antibiotic from *Streptomyces clavuligerus*. I. Discovery and biological activity. J. Antibiot. (Tokyo) 36:208-212.

Rappu, P., B.S. Shin, H. Zalkin, and P. Mantsala. 1999. A role for a highly conserved protein of unknown function in regulation of *Bacillus subtilis purA* by the purine repressor. J. Bacteriol. 181:3810-3815.

Reeder, T., and R. Schleif. 1991. Mapping, sequence, and apparent lack of function of

araJ, a gene of the *Escherichia coli* arabinose regulon. J. Bacteriol. 173:7765-7771.

Reznikoff, W.S. 2003. Tn*5* as a model for understanding DNA transposition. Mol. Microbiol. 47:1199-1206.

Rodriguez-Garcia, A., A. de La Fuente, R. Perez-Redondo, J.F. Martin, and P. Liras. 2000. Characterization and expression of the arginine biosynthesis gene cluster of *Streptomyces clavuligerus*. J. Mol. Microb. Biotechnol. 2:543-550.

Salowe, S.P., W.J. Krol, D. Iwata-Reuyl, and C.A. Townsend. 1991. Elucidation of the order of oxidations and identification of an intermediate in the multistep clavaminate synthase reaction. Biochemistry 30:2281-2292.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. Molecular cloning a laboratory manual. Second edition. Cold Spring Harbor Laboratory Press.

Samuel, S.J., S. Tzung, and S.A. Cohen. 1997. Hrp12, a novel heat-responsive, tissue-specific, phosphorylated protein isolated from mouse liver. Hepatology 25:1213-1222.

Smith, B., and P. Dyson. 1995. Inducible transposition in *Streptomyces lividans* of insertion sequence IS*6100* from *Mycobacterium fortuitum*. Mol. Microbiol. 18:933-941.

Snyder, L., and W. Champness. 1997. Transposition and nonhomologous recombination. p. 195-213. Molecular Genetics of Bacteria. American Society for Microbiology Washington D.C.

Sohaskey, C.D., H. Im, A.D. Nelson, and A.T. Schauer. 1992. Tn*4556* and luciferase: synergistic tools for visualizing transcription in *Streptomyces*. Gene 115:67-71.

Tahlan, K., A. Wong, P.H. Beatty, and S.E. Jensen. 2003. The paralogue gene cluster of *Streptomyces clavuligerus*: Isolation and characterization of a second set of genes encoding enzymes involved in the early stages of clavulanic acid and clavam metabolite biosynthesis. Manuscript.

Tang, L., Y.X. Zhang, and C.R. Hutchinson. 1994. Amino acid catabolism and antibiotic synthesis: valine is a source of precursors for macrolide biosynthesis in *Streptomyces ambofaciens* and *Streptomyces fradiae*. J. Bacteriol. 176:6107-6119.

Thai, W., A.S. Paradkar, and S.E. Jensen. 2001. Construction and analysis of Beta-lactamase-inhibitory protein (BLIP) non-producer mutants of *Streptomyces*

clavuligerus. Microbiology 147:325-335.

Trimble, J.J., S.C. Murthy, A. Bakker, R. Grassmann, and R.C. Desrosiers. 1988. A gene for dihydrofolate reductase in a herpesvirus. Science 239:1145-1147.

Ward, J.M., and J.E. Hodgson. 1993. The biosynthetic genes for clavulanic acid and cephamycin production occur as a 'super-cluster' in three *Streptomyces*. FEMS Microbiol. Lett. 110:239-242.

Weaden, J., and P. Dyson. 1998. Transposon mutagenesis with IS*6100* in the avermectin-producer *Streptomyces avermitilis*. Microbiology 144:1963-1970.

White, J., and M. Bibb. 1997. *bldA* dependence of undecylprodigiosin production in *Streptomyces coelicolor* A3(2) involves a pathway-specific regulatory cascade. J. Bacteriol. 27:627-633.

Wietzorrek, A., and M. Bibb. 1997. A novel family of proteins that regulates antibiotic production in streptomycetes appears to contain an OmpR-like DNA-binding fold. Mol. Microbiol. 25:1181-1184.

Wolkow, C.A., R.T. DeBoy, and N.L. Craig. 1996. Conjugating plasmids are preferred targets for Tn*7*. Genes Dev. 10:2145-2157.

Wright, F., and M.J. Bibb. 1992. Codon usage in the G + C-rich *Streptomyces* genome. Gene 113: 55-65.

Wu, J., and B.P. Rosen. 1991. The ArsR protein is a trans-acting regulatory protein. Mol. Microbiol. 5:1331-1136.

Wu, T.-K., R.W. Busby, T.A. Houston, D.B. McIlwaine, L.A. Egan, and G.A. Townsend. 1995. Identification, cloning, sequencing, and overexpression of the gene encoding proclavaminate amidino hydrolase and characterization of protein function in clavulanic acid biosynthesis. J. Bacteriol. 177:3714-3720.